

Universidade de Brasília – UnB
Faculdade UnB Gama – FGA
Engenharia de Software

Sistema para Auxiliar a Detecção da Doença de Parkinson Interpretando Sinais de Tremor de Repouso Parkinsoniano Utilizando Aprendizado de Máquinas

Autores: Jônntas Lennon Lima Costa e Flávio Costa Paixão
Orientadora: Dra. Lourdes Mattos Brasil
Coorientador: Bel. Ithallo Junior Alves Guimarães

Brasília, DF
2018



Jônnatas Lennon Lima Costa e Flávio Costa Paixão

**Sistema para Auxiliar a Detecção da Doença de
Parkinson Interpretando Sinais de Tremor de Repouso
Parkinsoniano Utilizando Aprendizado de Máquinas**

Monografia submetida ao curso de graduação
em Engenharia de Software da Universidade
de Brasília, como requisito parcial para ob-
tenção do Título de Bacharel em Engenharia
de Software.

Universidade de Brasília – UnB

Faculdade UnB Gama – FGA

Orientadora: Dra. Lourdes Mattos Brasil

Coorientador: Bel. Ithallo Junior Alves Guimarães

Brasília, DF

2018

Jônatas Lennon Lima Costa e Flávio Costa Paixão

Sistema para Auxiliar a Detecção da Doença de Parkinson Interpretando Sinais de Tremor de Repouso Parkinsoniano Utilizando Aprendizado de Máquinas/
Jônatas Lennon Lima Costa e Flávio Costa Paixão. – Brasília, DF, 2018-

102 p. : il. (algumas color.) ; 30 cm.

Orientadora: Dra. Lourdes Mattos Brasil

Trabalho de Conclusão de Curso – Universidade de Brasília – UnB
Faculdade UnB Gama – FGA , 2018.

1. Aprendizado de máquinas. 2. KNN. I. Dra. Lourdes Mattos Brasil. II. Universidade de Brasília. III. Faculdade UnB Gama. IV. Sistema para Auxiliar a Detecção da Doença de Parkinson Interpretando Sinais de Tremor de Repouso Parkinsoniano Utilizando Aprendizado de Máquinas

CDU 02:141:005.6

Jônnatas Lennon Lima Costa e Flávio Costa Paixão

Sistema para Auxiliar a Detecção da Doença de Parkinson Interpretando Sinais de Tremor de Repouso Parkinsoniano Utilizando Aprendizado de Máquinas

Monografia submetida ao curso de graduação em Engenharia de Software da Universidade de Brasília, como requisito parcial para obtenção do Título de Bacharel em Engenharia de Software.

Trabalho aprovado. Brasília, DF, 10 de Dezembro de 2018 – Data da aprovação do trabalho:

Dra. Lourdes Mattos Brasil

Orientadora

Bel. Ithallo Junior Alves Guimarães

Coorientador

MSc. Roberto Aguiar Lima

Convidado 1

Bel. Renata Menezes Lopes

Convidada 2

Brasília, DF

2018

*Este trabalho é dedicado à todos que trabalham duro com o
objetivo de mudar de vida e transformar o mundo.*

Agradecimentos

Aos nossos pais e amigos pelo apoio e incentivo durante todos estes anos.

A Universidade de Brasília, pelo curso de alto nível oferecido, que nos proporcionou experiências enriquecedoras.

Aos mestres Jorge Luiz Ferreira da Silva Júnior e Bruna da Silva Sousa, que nos deram suporte e compartilharam os conhecimentos necessários para tornar este projeto viável.

Ao nosso coorientador Ithallo Junior Alves Guimarães, por nos fornecer tutoria ao longo do projeto.

Agradeço a Dra. Lourdes Mattos Brasil, pela sua orientação e experiências acadêmicas compartilhadas.

Agradeço também aos demais membros da banca pela disponibilidade.

“Quando a caminhada fica dura, só os duros continuam caminhando.”
(Mano Brown)

Resumo

A Doença de Parkinson é uma doença neurológica que atinge uma grande parcela da população mundial, sendo majoritariamente pessoas idosas. Apesar de não terem sido descobertos os motivos efetivos que causam a doença, seu diagnóstico é possível graças aos sinais evidentes causados pela presença da mesma, sendo os mesmos suficientes. Ainda existe uma necessidade de se obter mais informações sobre o estado do paciente dessa doença para possíveis estudos sobre a mesma. Para garantir o acesso à essas informações e auxiliar no diagnóstico da Doença de Parkinson é possível utilizar os conceitos de Aprendizado de Máquina, que tem sido muito requisitado em diversas áreas do conhecimento, incluindo a área da saúde, e para a solução de problemas de diversos contextos. Este trabalho de conclusão de curso implementa um sistema denominado DPDP (Detector Preliminar da Doença de Parkinson) no qual é aplicado um modelo de Aprendizado de Máquina para o auxílio no diagnóstico preliminar da doença de Parkinson. Isso foi feito utilizando amostras recolhidas por dispositivos de coletas de sinais de eletromiografia de superfície (sEMG), afim de reconhecer padrões e obter novas informações sobre a doença em questão. O sistema utiliza o algoritmo *Random Forest* para detecção dos padrões da doença, pois o mesmo obteve maior sucesso em comparação a outros modelos estudados neste trabalho.

Palavras-chaves: Parkinson; Aprendizado de Máquina; sEMG.

Abstract

Parkinson's disease is a neurological disease that affects a large portion of the world's population, most of whom are elderly. Although the actual reasons that cause the disease have not been discovered, its diagnosis is possible thanks to the evident signs caused by the presence of the disease, it being sufficient. There is still a need to obtain more information about the state of the patient of this disease for possible studies about it. In order to guarantee access to this information and to assist in the diagnosis of Parkinson's Disease it is possible to use the concepts of Machine Learning, which has been widely requested in several areas of knowledge, including health, and to solve problems of several contexts. This final project implements a system called DPDP (Preliminary Detector of Parkinson's Disease, *Detector Preliminar da Doença de Parkinson*) in which a Machine Learning model is applied to aid in the preliminary diagnosis of Parkinson's disease. This was done using Surface Electromyography (sEMG) signals, in order to recognize patterns and to obtain new information about the disease in question. The system uses the algorithm Random Forest to detect the patterns of the disease, once it was more successful in comparison with other models studied in this work.

Key-words: Parkinson; Machine Learning; sEMG.

Lista de ilustrações

Figura 1 – Ilustração da matriz de confusão, adaptado de (PEDREGOSA et al., 2011).	33
Figura 2 – Ilustração do <i>k-fold</i> , adaptado de Raschka (2015).	35
Figura 3 – SVMs utilizando <i>kernel</i> diferentes, adaptado de Pedregosa et al. (2011).	36
Figura 4 – Esquematização de um SVM, adaptado de Pedregosa et al. (2011).	37
Figura 5 – Tesselação de Voronoi, adaptado de Voronoi (1907).	42
Figura 6 – Diagrama da substância negra, adaptado de Maia (2018).	43
Figura 7 – Modelo de uma unidade motora, adaptado de Sabino (2017).	47
Figura 8 – Sinal sEMG muscular, adaptado de Ortolan (2002).	47
Figura 9 – <i>Features</i> de um sEMG, adaptado de Yousefi e Hamilton-Wright (2014).	51
Figura 10 – Diagrama do processo (Imagem dos autores).	55
Figura 11 – Amostra de um sinal após a aplicação do FFT no canal 1 (Imagem dos autores).	60
Figura 12 – Amostra de um sinal bruto no canal 1 (Imagem dos autores).	61
Figura 13 – Comparativo entre os canais 1 e 2, com a FFT (Imagem dos autores).	61
Figura 14 – Comparativo entre os canais 3 e 4, com a FFT (Imagem dos autores).	62
Figura 15 – Matriz de confusão com o filtro passa-banda e sem o filtro no Ch1 (Imagem dos autores).	62
Figura 16 – Pagina inicial (Imagem dos autores).	64
Figura 17 – Cadastro das pacientes.	65
Figura 18 – Editar dados dos pacientes (Imagem dos autores).	66
Figura 19 – Avaliar pacientes (Imagem dos autores).	67
Figura 20 – Pesquisar itens pelas colunas (Imagem dos autores).	67
Figura 21 – <i>Download</i> da tabela, impressão e exibir colunas (Imagem dos autores).	68
Figura 22 – <i>Download</i> da tabela, impressão e exibir colunas (Imagem dos autores).	69
Figura 23 – informações sobre o projeto, na página Sobre (Imagem dos autores).	70
Figura 24 – Diagrama de pacotes.	80
Figura 25 – Diagrama de Classes.	82

Lista de tabelas

Tabela 1 – Comparativo da acurácia entre os canais com a FFT.	60
Tabela 2 – Comparativo entre a acurácia dos algoritmos do Ch1	63
Tabela 3 – Instrução do Problema	84
Tabela 4 – Resumo da Parte Interessada	84
Tabela 5 – Resumo do Usuário	85
Tabela 6 – Perfil dos Orientandos	85
Tabela 7 – Perfil dos Orientadores	85
Tabela 8 – Perfil dos Médicos	86
Tabela 9 – Resumo dos recursos	86
Tabela 10 – Requisitos Funcionais	87

Lista de abreviaturas e siglas

AM	Aprendizado de máquina
DP	Doença de Parkinson
sEMG	Eletromiografia de Superfície (sEMG, do inglês: <i>Surface Electromyography</i>)
SVM	Máquina de vetores de suporte (SVM, do inglês: <i>support vector machine</i>)
KNN	K vizinhos mais próximos (KNN, do inglês: <i>k-nearest neighbors</i>)
API	Interface de Programação de Aplicação (API, do Inglês: <i>Application Programming Interface</i>)
ATP	Adenosina Trifosfato
ACo	Acetilcolina
UM	Unidade motora
EMD	Decomposição em modo empírico (EMD, do inglês: <i>empirical mode decomposition</i>)
FFT	Transformada rápida de Fourier (FFT, do inglês: <i>fast Fourier transform</i>)
IMF	Funções do modo intrínseco (IMF, do inglês: <i>intrinsic mode functions</i>)
SVD	Decomposição em valores singulares (SVD, do inglês: <i>singular value decomposition</i>)
DWT	Transformada discreta de <i>Wavelets</i> (do inglês: <i>discrete wavelet transform</i>)
BPNM	Modelo e Notação de Processos de Negócio (BPNM, do Inglês: <i>Business Process Model and Notation</i>)
EHR	Espaço Hilbert de Reprodução
RBF	Função de base radial (RBF, do inglês: <i>radial basis function</i>)
RNA	Rede neural artificial

PEPS	Processo Pós-Sináptico
BPMN	Notação de Processos de Negócio (BPMN, do inglês: <i>Business Process Model and Notation</i>)

Sumário

1	INTRODUÇÃO	25
1.1	Preâmbulo	25
1.2	Justificativa	25
1.3	Objetivos	26
1.3.1	Objetivo Geral	26
1.3.2	Objetivos Específicos	26
2	APRENDIZADO DE MÁQUINA	27
2.1	Definição	27
2.2	Aplicações em Aprendizado de Máquina	27
2.3	Funcionamento de um algoritmo de AM	27
2.4	Tipos de Aprendizado	28
2.4.1	Aprendizado Supervisionado	28
2.4.2	Aprendizado não-supervisionado	29
2.4.3	Aprendizado por reforço	29
2.5	Feature	29
2.5.1	PCA	30
2.5.2	FFT	30
2.6	Desafios do AM	30
2.7	Ferramentas	31
2.8	Visualizar a solução	32
2.8.1	Matriz de Confusão	33
2.9	Validação cruzada	33
2.9.1	<i>Holdout Cross-Validation</i>	34
2.9.2	<i>k-Fold Cross-Validation</i>	34
2.9.3	<i>Leave-One-Out Cross-Validation (LOOCV)</i>	34
2.10	SVM	35
2.10.1	Funcionamento	36
2.11	Random Forest	39
2.11.1	Funcionamento	39
2.12	KNN	40
2.12.1	Funcionamento	41
3	A DOENÇA DE PARKINSON	43
3.1	Definição	43
3.2	Diagnóstico e Sintomas	44

3.2.1	Tratamento	44
4	ELETROMIOGRAFIA DE SUPERFÍCIE	45
4.1	Definição	45
4.2	Características do sinal	45
4.2.1	Contração muscular	45
4.2.2	Unidade Motora	46
4.3	Aplicações da sEMG	47
5	TRABALHOS RELACIONADOS	49
5.1	AM, DP e sEMG	49
5.1.1	Análise	49
5.1.2	Decomposição	49
6	SOLUÇÃO DE <i>SOFTWARE</i>	53
6.1	Ágil	53
6.1.1	<i>Scrum</i>	53
6.2	Ferramentas	53
6.2.1	Scikit-learn	54
6.2.2	Django-Rest	54
6.2.3	React	54
6.3	Produto Mínimo Viável	54
7	METODOLOGIA	55
7.1	Comitê de Ética e Pesquisa	55
7.2	Processo de construção de uma aplicação de Aprendizado de Máquina	55
7.2.1	Abstrair o problema no contexto	56
7.2.2	Coleta de dados	56
7.2.3	Analisar os Dados	56
7.2.4	Processar Dados	57
7.2.5	Selecionar o Modelo	58
7.2.6	Treinar o Modelo	58
7.2.7	Validar a solução	58
7.2.8	Apresentar a solução	58
8	RESULTADOS	59
8.1	Processo de AM	59
8.1.1	Coleta	59
8.1.2	Processamento dos dados	59
8.2	Solução de <i>software</i>	63
8.2.1	Arquitetura	63

8.2.2	Visão	63
8.2.3	Produto	63
8.2.3.1	Pagina inicial	64
8.2.3.2	Cadastro de pacientes	64
8.2.3.3	Edição de pacientes	65
8.2.3.4	Avaliar pacientes	65
8.2.3.5	Pesquisar pacientes	65
8.2.3.6	Pesquisar pacientes	66
8.2.3.7	Pagina arquivados	66
8.2.3.8	Página sobre	66
8.3	Passos futuros e considerações finais	66

REFERÊNCIAS	71
------------------------------	-----------

APÊNDICES	77
----------------------------	-----------

	APÊNDICE A – DOCUMENTO DE ARQUITETURA	79
A.1	Introdução	79
A.2	Tecnologias	79
A.2.1	SPA	79
A.2.2	REST	79
A.3	Representação da Arquitetura	79
A.4	Visão Lógica	80
A.4.1	Diagrama de Pacotes	80
A.4.1.1	<i>Machine Learning models</i>	80
A.4.1.2	API	81
A.4.1.3	Web	81
A.4.2	Diagrama de Classes	81
A.5	Restrições e Metas Arquiteturais	81
A.5.1	Restrições	81
	APÊNDICE B – DOCUMENTO DE VISÃO	83
B.1	Introdução	83
B.1.1	Problema	83
B.1.2	Escopo	83
B.2	Posicionando	83
B.2.1	Oportunidade de Negócios	83
B.2.2	Instrução do Problema	84
B.3	Descrições da Parte Interessada e do Usuário	84

B.3.1	Resumo da Parte Interessada	84
B.3.2	Resumo do Usuário	84
B.3.3	Perfis das Partes Interessadas	85
B.3.3.1	Orientandos	85
B.3.3.2	Orientadores	85
B.3.4	Perfis do Usuário	85
B.3.4.1	Médicos	86
B.4	Descrição da Solução	86
B.4.1	Perspectiva do Produto	86
B.4.2	Resumo dos recursos	86
B.4.3	Recursos do Produto	86
B.4.4	Requisitos Funcionais	87

ANEXOS 89

ANEXO A – PROTOCOLO - GRUPO DE PESQUISA SOBRE COM- PLICAÇÕES DA DOENÇA DE PARKINSON 91

ANEXO B – TAE 101

1 Introdução

1.1 Preâmbulo

Este trabalho visa o desenvolvimento de um *software* capaz de auxiliar o diagnóstico da DP (Doença de Parkinson) avaliando tremores patológicos de repouso, utilizando AM (Aprendizado de Máquina) para identificar se é uma pessoa com deficiência ou não. Com relação aos dados que serão analisados pelo algoritmo, esses serão coletados através da sEMG (sigla do inglês: *surface electromyography*, eletromiografia de superfície).

A DP é uma doença crônica que afeta o sistema nervoso central, prejudicando, entre outros fatores, principalmente a mobilidade dos doentes. A doença não possui cura e sua causa ainda é desconhecida. Deste modo, o tratamento é focado na melhora de qualidade de vida dos pacientes (COSTA; DIAFÉRIA; BEHLAU, 2016). O diagnóstico da doença é clínico, onde identifica-se a presença de três principais sintomas a bradicinesia, os tremores corporais em repouso e a rigidez corporal. Atualmente, existem estudos que buscam identificar a DP utilizando sEMG (EFTAXIAS et al., 2015).

A sEMG é uma técnica rápida e não invasiva para monitorar a atividade elétrica envolvendo as fibras musculares, através do uso de eletrodos sobre a pele do paciente, gerando um sinal eletromiográfico. Um sinal eletromiográfico é o somatório das contrações musculares identificadas pelos eletrodos (PERNAMBUCO et al., 2010). Este sinal gerado não é de fácil compreensão e pode ser melhor analisado utilizando AM, como realizado por Kugler et al. (2013) e Loconsole et al. (2018).

A AM é a uma área da inteligência artificial e da ciência de dados, onde se estuda o desenvolvimento de algoritmos autônomos capazes de aprender com os seus erros e realizar previsões sobre novos dados. Se executado corretamente obtém um alto nível de acerto, muitas vezes superior ao realizado por humanos (KOHAVI, 1998).

1.2 Justificativa

Atualmente, o diagnóstico da DP é basicamente clínico, onde o neurologista identifica alguns sintomas associados a doença, como a bradicinesia, tremor em repouso, rigidez muscular e instabilidade postural (GAGO et al., 2014). A sEMG pode auxiliar esse processo, porém a onda gerada não é de fácil interpretação, sendo que com AM é possível obter um resultado rápido e com alta precisão, como demonstrado por Camara et al. (2015), onde obteve-se aproximadamente 98% de acerto num possível diagnóstico da doença.

Logo, com o desenvolvimento do *software* proposto neste trabalho, este processo

de diagnóstico poderá ser facilitado utilizando um único exame não invasivo de sEMG, combinado com AM para classificar os deficientes. Assim auxiliando com rapidez e eficácia a detecção da doença.

1.3 Objetivos

1.3.1 Objetivo Geral

O Objetivo deste Trabalho é o desenvolvimento de um *software* utilizando algoritmos de AM capazes de auxiliar no diagnóstico da DP, focando em analisar o estado do tremor de repouso do paciente, classificando os indivíduos em deficientes da DP e não deficientes.

1.3.2 Objetivos Específicos

- Analisar o comportamento dos algoritmos SVM (sigla do inglês: *Support Vector Machine*, Máquina de vetores de suporte), KNN (sigla do inglês: *k-nearest neighbors*, K-vizinhos mais próximos) e *Random Forest*;
- Definir modelo a ser utilizado;
- Definir a arquitetura do *software*;
- Desenvolver um sistema integrando o algoritmo treinado anteriormente, com um *software* amigável ao usuário;

2 Aprendizado de Máquina

2.1 Definição

A AM é considerado uma disciplina dentro da inteligência artificial, com foco na construção de softwares computacionais dotados da capacidade de aprender autonomamente (HOSCH, 2009). Essa disciplina trabalha com o estudo e a construção de algoritmos, capazes de aprender com seus “erros” no reconhecimento de dados e realizarem previsões sobre novos dados. Esses algoritmos de indução são considerados um grande passo na descoberta do conhecimento (KOHAVI, 1998).

Apesar da complexidade em AM, essa disciplina pode ser contextualizada de maneira mais geral como o campo de estudo que dá ao computador a habilidade de aprender sem ser explicitamente programado (SAMUEL, 1959), porém uma visão mais orientada para a área da engenharia, que segundo Mitchell et al. (1997), é dito que um programa de computador aprende com a experiência E em relação a alguma tarefa T e alguma medida de desempenho P , se seu desempenho em T é medido por P melhora com a experiência E .

2.2 Aplicações em Aprendizado de Máquina

O AM serve de auxílio para diversos contextos de naturezas diferentes. Sendo assim, a AM consegue auxiliar com o uso dos dados coletados a melhor decisão de um negócio, por exemplo, uma possível aplicação de um modelo de AM é prever qual a probabilidade de um cliente comprar um determinado produto com base no seu histórico de compras passadas (AMAZON, 2018).

Na saúde, AM é uma tendência crescente na indústria médica, graças ao advento de dispositivos eletrônicos como sensores, que podem acessar os dados de pacientes em tempo real. A tecnologia também pode ajudar médicos a analisar os dados para identificar tendências ou alertas, consequentemente, levando ao aperfeiçoamento de diagnósticos e tratamentos (SAS, 2018).

2.3 Funcionamento de um algoritmo de AM

Cada algoritmo tem suas peculiaridades quanto ao seu funcionamento. Entretanto esses algoritmos em seu estado mais básico trabalham com a entrada de dados num sistema, que processa os mesmos utilizando um modelo de predição geralmente já postulado,

no qual derivam um resultado com base em uma predição que o modelo fez sobre os dados de entrada. Esses algoritmos possuem alguns métodos e técnicas para validar seu funcionamento, por exemplo, declarando uma dada porcentagem para avaliar o grau da acurácia de seu resultado, permitindo assim verificar se a predição gerada pelo algoritmo se aproxima do resultado esperado no contexto real. Sendo essa e outras situações melhor contextualizado neste capítulo.

Considerando ainda que o AM, com seu foco na predição no desenvolvimento de algoritmos capazes de aprender com seus erros, essa disciplina é muito atrelada a otimização matemática. Possuindo como uma de suas principais características, a utilização de métodos estatísticos para validar o seu funcionamento (TREVOR; ROBERT; JH, 2009).

2.4 Tipos de Aprendizado

Os algoritmos são classificados quanto ao seu tipo de aprendizado. Existem três categorias, as quais se distinguem de acordo com a natureza do comportamento no tratamento dos dados. Estas categorias são aprendizado supervisionado, não supervisionado e por reforço (GÉRON, 2017).

2.4.1 Aprendizado Supervisionado

O objetivo principal do aprendizado supervisionado é ensinar um modelo através de uma base de treino etiquetada, que permite fazer predições de dados futuros. O termo supervisionado se refere a um grupo de dados onde a “resposta”, sinal ou valor esperado já é conhecido. Este valor é conhecido como etiqueta e é utilizada para verificar se o algoritmo identificou corretamente os dados durante o treinamento. As tarefas do aprendizado supervisionado podem ser divididas em classificação e regressão, onde a classificação tem as etiquetas esperadas com um valor fixo, por exemplo, uma avaliação binária, e a regressão é referente a um sinal com valor contínuo, como uma regressão polinomial (GÉRON, 2017), ou segundo Kirk (2014), a regressão é ajustar o conjunto de dados á uma função qualquer.

Exemplos de algoritmos existentes (LORENA; CARVALHO, 2007); (KIRK, 2014); (MURTHY, 1998); (HO, 1995):

- **KNN**: determina o rótulo de classificação de um dado baseando nos valores das amostras vizinhas mais próximos;
- **Decision Trees**: uma forma de organizar regras referentes a dados com estruturas hierárquicas e sequências, que particionam estes dados recursivamente;

- **SVM**: separado em linear e não linear, consiste na separação de dados de um conjunto maior em classes de conjuntos mais restritos.
- **Random Forest**: método de aprendizado conjunto para classificação, regressão e outras tarefas, que operam construindo uma multiplicidade de árvores de decisão no momento do treinamento e gerando a classe que é o modo das classes ou previsão média das árvores individuais.

2.4.2 Aprendizado não-supervisionado

Ao contrário do aprendizado supervisionado, os dados não estão etiquetados ou até mesmo possuem uma estrutura desconhecida. Utilizando as técnicas do aprendizado não supervisionado é possível explorar a estrutura para extrair informações necessárias dos dados sem a orientação de uma variável de resultado conhecida ou uma função de recompensa (GÉRON, 2017).

Exemplos de algoritmos existentes (KIRK, 2014):

- **Clustering**: É utilizado para agrupar padrões, ou melhor, no treinamento é introduzido um conjunto de dados e o algoritmo procura relações entre eles agrupando tais dados em grupos por grau de semelhança, na predição, com um novo dado inserido o algoritmo busca padrões existentes nele e cataloga-o no respectivo grupo;
- **Collaborative Filtering**: É utilizado no desenvolvimento de sistemas de recomendação, basicamente utiliza informações cruzadas dos usuários para predizer gostos em comum, por exemplo se você compra bastante cerveja em um site de compras, o sistema recomendará cerveja para você, caso um grupo significativo de usuários compre cerveja e carvão, o sistema recomendará conjuntamente o carvão.

2.4.3 Aprendizado por reforço

Utiliza um sistema de recompensas para reforçar a continuidade de um comportamento esperado, por exemplo, um jogo onde o personagem é beneficiado com uma recompensa sempre que atinge um golpe no adversário, quanto mais forte for o golpe maior a recompensa. Deste modo, o modelo memoriza quais os comandos geraram as maiores recompensas (KIRK, 2014).

2.5 Feature

Uma *feature* é uma característica ou variável do problema, que pode ser processada pelo modelo e gerar um resultado condizente com a realidade. Deve ser mensurável e independente, sendo que uma *feature* é a base para o funcionamento dos algoritmos de

classificação, regressão e reconhecimento de padrões. Normalmente, é um número mas pode ser uma palavra, grafo, vetor, dentre outros. (CHANDRASHEKAR; SAHIN, 2014).

Outro conceito importante sobre AMs é a **redução de dimensionalidade**, muito utilizada em estatística, refere-se ao processo de diminuição do número de variáveis aleatórias, selecionando um subconjunto com variáveis relevantes. Serve, dentre outras funcionalidades, para transformar o processo de aprendizado mais eficiente (BORGES; NIEVOLA, 2006). Isto ocorre devido a modelos com um número elevado de variáveis ocasionam em um maior tempo de computação e um baixo desempenho da previsão (CHANDRASHEKAR; SAHIN, 2014).

Um exemplo de técnica de redução de dimensionalidade é a PCA (sigla do inglês: *Principal Component Analysis*, Análise de componentes principais), sendo esta uma das técnicas mais difundidas na redução de dimensão.

2.5.1 PCA

A PCA possibilita reduzir consideravelmente o número de variáveis disponíveis, além de, eliminar sobreposições e escolher formas mais representativas do conjunto de dados original, por meio de combinações lineares ortogonais das variáveis originais, para um novo sistema de coordenadas (JOLLIFFE, 2011).

2.5.2 FFT

A FFT (sigla do inglês: *Fast Fourier Transform*, Transformada Rápida de Fourier) é um algoritmo capaz de realizar a Transformada de discreta de Fourier e sua inversa. Este processo é realizado com uma alta eficiência, reduzindo a complexidade de $O(N^2)$ para $O(N * \log_2(N))$, ou seja de reduzindo o tempo de processamento em centenas de horas (BRIGHAM; BRIGHAM, 1988).

2.6 Desafios do AM

Em um projeto de AM, existem diversos aspectos que podem levar a um péssimo resultado ou problemas na execução do mesmo. Em resumo, a tarefa principal do AM é a seleção de um algoritmo, ou modelo, e treiná-lo utilizando algum tipo de dado. Logo, ao analisar os problemas provenientes de um resultado ineficiente, pode-se encontrar principalmente duas origens para o problema a escolha de um “algoritmo ruim” ou “dados ruins” (GÉRON, 2017).

O conceito de “dados ruins”, não deve ser considerado ao pé da letra, sendo um pouco mais complexo, envolvendo principalmente a coleta destes e tratamento associado. Para exemplificar, pode-se listar alguns exemplos retirados de Géron (2017):

- **Quantidade insuficiente de dados de treino:** Para um funcionamento pleno de um sistema de AMs é necessária uma grande quantidade de dados, alguns problemas necessitam de uma alta quantidade de dados para funcionar corretamente, como no reconhecimento de imagens.
- **Dados não representativos:** Com o intuito de alcançar a boa generalização, é crucial que os dados de treino sejam representativos em relação ao escopo dos dados.
- **Dados de baixa qualidade:** É evidente que caso os dados de treinos tenham muitos erros e ruídos, originados de um péssimo processo de coleta, isso irá dificultar o sistema na detecção de padrões necessários ou até mesmo alcançar o mínimo de performance requerido.
- **Features irrelevantes:** Seguindo a lógica por trás da expressão *garbage-in garbage-out* (termo tradução livre: Lixo entra e lixo sai), o sistema apenas será capaz de gerar bons resultados caso exista dados contendo o suficiente de *features* relevantes ao invés das não relevantes, ou seja, *features* irrelevantes podem conduzir o programa a uma baixa precisão.

Com relação a “algoritmo ruim”, também não deve ser interpretado literalmente, ou seja, um “algoritmo ruim” pode ser proveniente da escolha de um modelo de algoritmo para tratar um problema no qual o algoritmo não foi otimizado para solucionar. Existem dois principais problemas associados ao manejo de um algoritmo, sendo eles (GÉRON, 2017):

- **Overfitting dos dados de treino:** *Overfitting* acontece quando o modelo apresenta uma boa performance na sua base de treino, mas não generaliza muito bem, ou seja, o modelo pode ser considerado “viciado”, é um erro bem comum ocorrendo no treinamento do modelo. Por exemplo, uma amostragem com muito ruído pode encontrar algum padrão no próprio ruído e generalizar para o mesmo, não generalizando para novos dados.
- **Underfitting dos dados de treino:** Ao contrário do *Overfitting*, *Underfitting* acontece quando o modelo é muito simples para ser utilizado com a estrutura de dados selecionada, resultando em um modelo impreciso para ser utilizado no mundo real. Isto devido aos dados reais serem provavelmente mais complexos que os utilizados no treinamento do modelo.

2.7 Ferramentas

Existem diversas ferramentas para auxiliar o desenvolvimento em AM, sendo que atualmente *Python* é uma das linguagens mais utilizadas nesse segmento. Segundo Sen-

ders et al. (2017), essa utilização é devido ao conjunto bem eficiente de bibliotecas disponíveis, o fato de ser de código aberto, a facilidade de plotar gráficos e demais esquemas de visualização, além da simplicidade de identificar erros no *software*, tornam a linguagem *Python* uma boa indicação para trabalhar com AM.

Algumas ferramentas amplamente utilizadas em AM, podem ser elucidadas a seguir (PEDREGOSA et al., 2011); (ABADI et al., 2016); (MCKINNEY, 2012):

- **Scikit-learn**: Possui uma ampla gama de algoritmos, com uma alta facilidade de uso, boa documentação e uma API (sigla do Inglês: *Application Programming Interface*, Interface de Programação de Aplicação) amplamente consistente. API é um conjunto de funções que são utilizados para facilitar o desenvolvimento de um *software*;++
- **TensorFlow**: Projetado pelo Google para operar em larga escala, utiliza ferramentas avançadas sendo utilizado inclusive pelo próprio Google;
- **Keras**: API de alto nível com foco em RNAs (Redes Neurais Artificiais);
- **Matplotlib**: Ferramenta para a criação de gráficos;
- **Numpy**: Utilizada para manipulações matemáticas complexas, como multiplicação de matrizes, entre outras;
- **Jupyter Notebook**: aplicativo que permite a criação de códigos em *Python* de modo interativo;
- **Pandas**: Ferramenta para estruturar dados, tornando o trabalho com dados “relacionais” ou “rotulados” mais fácil e intuitivo. Ideal para trabalhar com matrizes.

2.8 Visualizar a solução

Construir visualizações informativas são umas das mais importantes tarefas dentro da análise de dados (MCKINNEY, 2012). Essa etapa consiste na utilização de ferramentas e técnicas para a representação dos dados obtidos tanto pelo modelo quanto pelos dados coletados para validação dos mesmos. O resultados obtidos pelo algoritmo de AM sobre os dados de coleta serem aferidos quanto ao seu *benchmarking* (CAMP, 1998) para avaliar o seu desempenho, análise de acurácia e precisão como também para avaliar seu tempo de resposta. A apresentação dos dados da coleta tem por objetivo avaliar tanto a separação do espaço amostral dos dados quanto para se obter um comparativo que melhor represente de maneira visual os dados após a filtragem da etapa do processamento dos dados, de maneira geral serve para ajudar a identificar *outliers* ou transformações de dados necessárias ou como uma forma de gerar ideias para modelos (MCKINNEY, 2012).

2.8.1 Matriz de Confusão

Um dos métodos mais utilizados para apresentar a solução do modelo é a matriz de confusão. Uma matriz de confusão é um método para visualizar graficamente a eficiência de um modelo de AM. Onde, realiza-se o cruzamento entre os valores esperados com os valores alcançados pelo algoritmo (CAELEN, 2017). Por exemplo na Figura 1, existem três tipos de flores e o algoritmo, no caso o SVM, deveria classificar os dados de entradas em um destes três grupos. Como observado, existiam 13 dados do tipo “setosa”, e o SVM acertou todos os 13. O mesmo ocorre com os dados do tipo “virginica”, onde ocorreu 100% de acerto. Porém, nos dados do tipo “versicolor” existiam 16 amostras, e o SVM acertou 10 e errou os outros 6 restantes, já que o valor alcançado nestes 6 dados foi de “virginica” (PEDREGOSA et al., 2011).

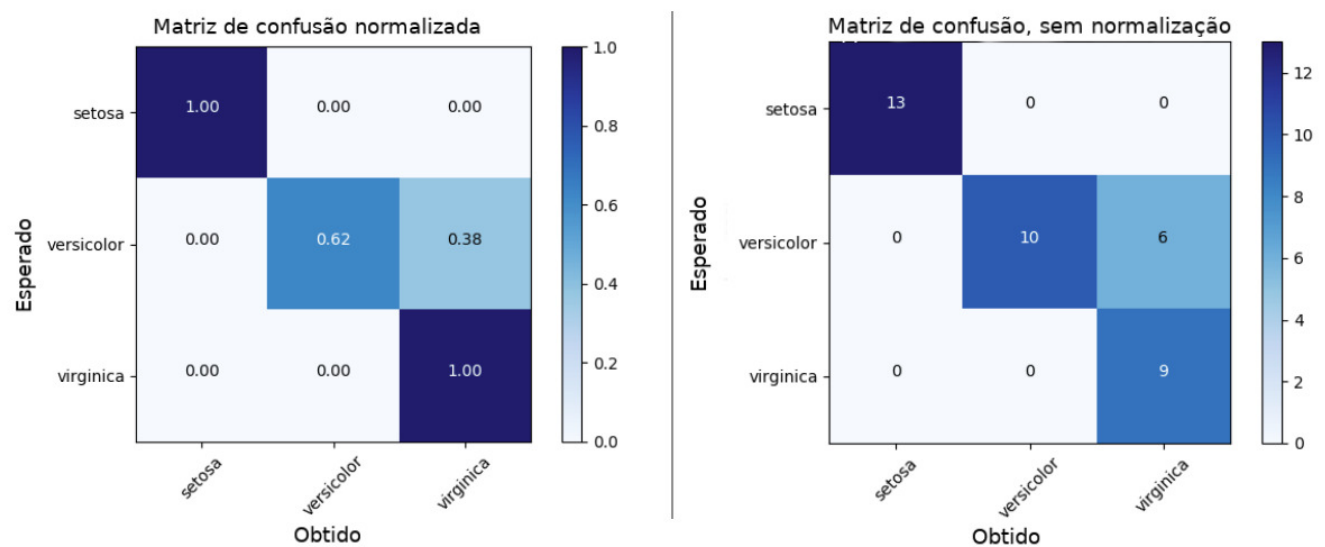


Figura 1 – Ilustração da matriz de confusão, adaptado de (PEDREGOSA et al., 2011).

2.9 Validação cruzada

A validação cruzada ou *Cross-Validation* é uma técnica para verificar o quão eficiente é a capacidade de generalização do modelo. Ou seja, qual o desempenho do modelo ao prever um novo conjunto de dados. Com a validação cruzada evita-se o *overfitting*, além de auxiliar no ajuste dos parâmetros do algoritmo (JAMES et al., 2013). A validação cruzada consiste em (JAMES et al., 2013) e (RASCHKA, 2015):

1. Separar aleatoriamente o conjunto de dados em três partes (treino, teste e validação);
2. O conjunto de dados de treino é utilizado para treinar o modelo;
3. O conjunto de validação é utilizado para ajustar os parâmetros. Isto é, após o treino verifica-se a precisão do modelo e caso não atinga a expectativa, realiza-se os ajustes

necessários no algoritmo. Após, realiza-se um novo treinamento com os dados de teste;

4. Ao final, verifica-se a generalização do algoritmo treinado, utilizando o conjunto de validação. Deste modo, o conjunto de validação é utilizada apenas uma vez no processo de AM.

A validação cruzada possui ainda algumas variações, sendo elas: *holdout cross-validation*, *k-Fold Cross-Validation* e *Leave-one-out cross-validation (LOOCV)* (RASCHKA, 2015).

2.9.1 *Holdout Cross-Validation*

Modelo clássico da validação cruzada, consiste em separar os dados em dois conjuntos mutualmente exclusivos (treino e validação), porém essa abordagem acaba não sendo eficiente, por aumentar a probabilidade de *overfitting*, para uma maior eficiência, utiliza-se a variações com os três conjuntos explicados anteriormente (RASCHKA, 2015).

2.9.2 *k-Fold Cross-Validation*

No *k-fold* divide-se o conjunto total de dados em k subconjuntos mutualmente exclusivos, com a mesma quantidade de dados, um destes subconjuntos é utilizado para teste e os outros $k-1$ restantes são utilizados no treinamento do modelo com os devidos ajuste de parâmetros necessários, tal processo é repetido k vezes, alternando o conjunto de teste nestas iterações. Após as k iterações, se calcula a média do erro esperado. A Figura 2 ilustra este método (RASCHKA, 2015).

2.9.3 *Leave-One-Out Cross-Validation (LOOCV)*

Uma variação do *k-fold*, com k igual ao total de dados de entrada N . Possui um alto custo operacional (RASCHKA, 2015).

As seções seguintes descrevem os principais algoritmos de AM utilizados em classificação, segundo Pedregosa et al. (2011), sendo eles:

- SVM;
- *Random forest*;
- KNN.

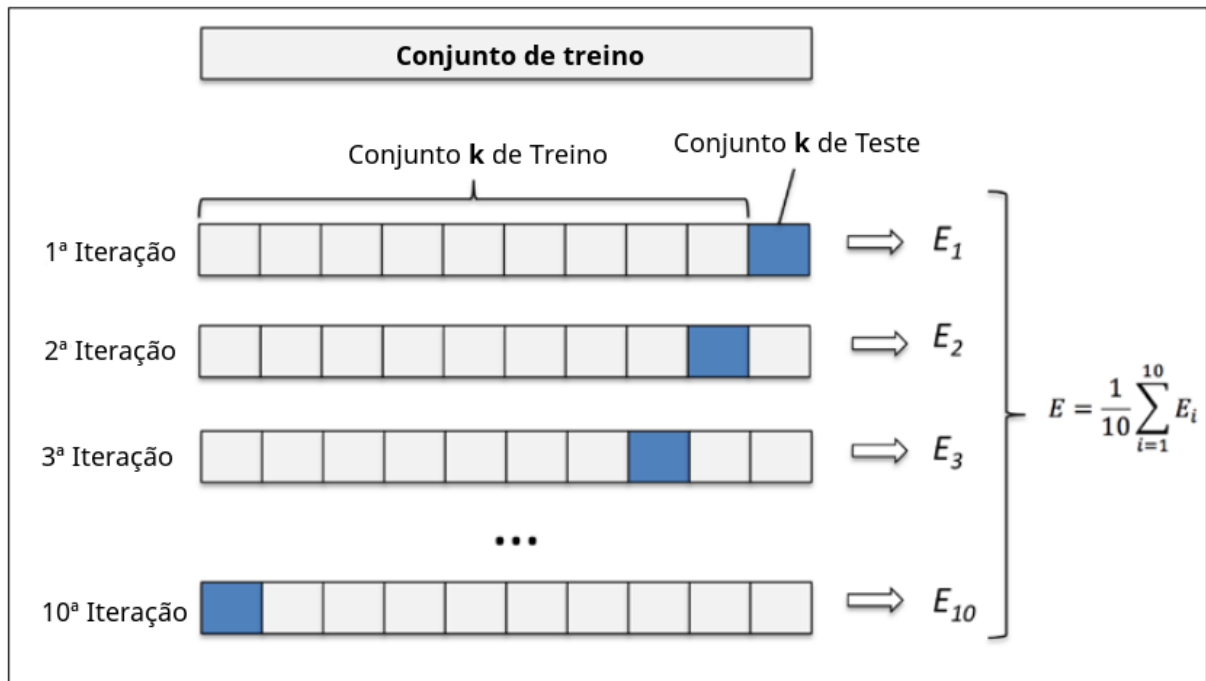


Figura 2 – Ilustração do k -fold, adaptado de Raschka (2015).

2.10 SVM

A SVM é um dos modelos mais populares em AMs, sendo composto por um conjunto de métodos de aprendizado supervisionada. Pode ser utilizada para classificação, regressão ou *outlier detection* (PEDREGOSA et al., 2011). Segundo Braga, Carvalho e Ludermir (2000) e Haykin (1999), os resultados da aplicação de SVM são comparáveis e muitas vezes superiores aos obtidos por outros algoritmos de aprendizado, como as RNAs. Exemplos de aplicações de sucesso podem ser encontrados em diversos domínios, como na categorização de textos, catalogação de imagens (PONTIL; VERRI, 1998), e em Bioinformática (NOBLE et al., 2004).

As SVMs são divididas em dois grupos, categorizadas conforme o estado dados a serem resolvidos, ou seja é um problema linear ou não (LORENA; CARVALHO, 2007).

Segundo Pedregosa et al. (2011), as principais vantagens em SVMs são:

- Sua eficácia em grandes espaços dimensionais ou quando o número de dimensões supera o de amostras;
- A eficiência em termo de memória;
- A versatilidade: possui diferentes tipos de funções de *kernel* (linear, polinomial, sigmoide e RBF (sigla do inglês: *radial basis function*, função de base radial), possibilita também a utilização de um *kernel* customizado, ou seja, uma função matemática

classificação personalizada. A Figura 3 exemplifica algumas funções de *kernel*, onde os dados são classificados em três grupos distintos.

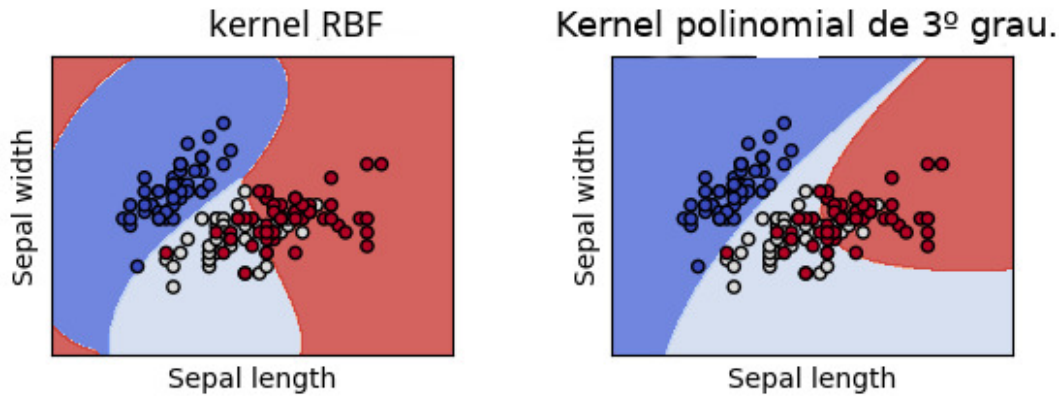


Figura 3 – SVMs utilizando *kernel* diferentes, adaptado de Pedregosa et al. (2011).

A SVM funciona no princípio de separação ótima entre classes. Ou seja, com base nas amostras de treinamento, cria-se uma distribuição de probabilidade, remove-se algumas amostras de modo iterativo e calcula-se novamente a distribuição de probabilidade, continuando este ciclo até estabelecer os vetores de suporte em um hiperplano ou em um conjunto de hiperplanos. Estes vetores de suporte, são os pontos que definem a maior margem possível entre as classes (HUANG; DAVIS; TOWNSHEND, 2002). A Figura 4 exemplifica estes conceitos.

2.10.1 Funcionamento

As SVMs foram desenvolvidas por Vapnik (1998), no âmbito da Teoria Estatística da Aprendizagem, logo para entender o funcionamento do dado modelo é necessário entender o próprio *framework* da Teoria Estatística da Aprendizagem (VAPNIK, 1998). Porém como o foco neste trabalho é a aplicação do modelo, não será abordado de maneira aprofundada o conceito da Teoria Estatística da Aprendizagem, cujo o foco é a implementação.

Em sua formulação mais geral, a SVM encontra um hiperplano em um espaço diferente daquele dos dados de entrada \mathbf{X} . É um hiperplano, em um espaço de característica induzido por um *kernel* \mathbf{K} . Através do *kernel* \mathbf{K} , o espaço de hipóteses é definido como um conjunto de “hiperplanos” no espaço de feições induzido por \mathbf{K} . Isso também pode ser visto como um conjunto de funções em um EHR (Espaço Hilbert de Reprodução) definido por \mathbf{K} (WAHBA, 1990) (VAPNIK, 1998). Para resumir, o espaço de hipóteses usado pela SVM é um subconjunto do conjunto de hiperplanos definido em algum espaço de um EHR. Este espaço pode ser formalmente escrito como na Equação 2.1.

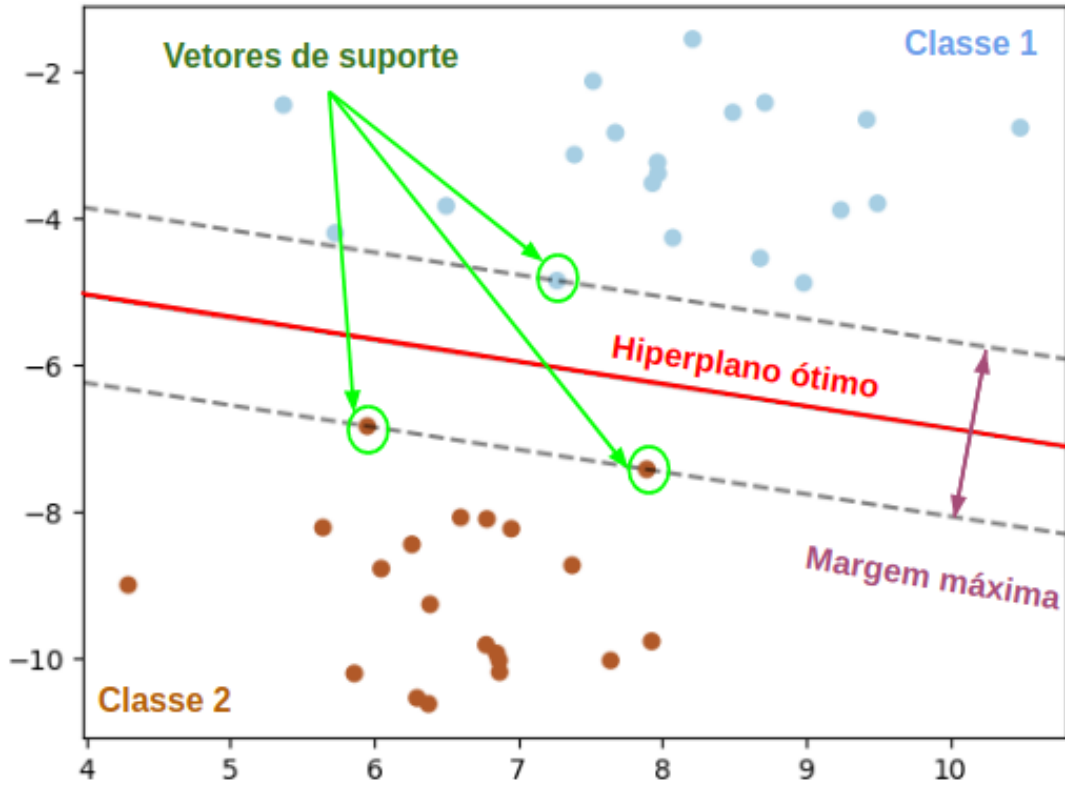


Figura 4 – Esquemática de um SVM, adaptado de [Pedregosa et al. \(2011\)](#).

$$\{f : \|f\|_2^k < \infty\} \quad (2.1)$$

Onde \mathbf{K} é o kernel que define o EHR, e $\|f\|_2^k$ é a norma EHR da função ([WAHBA, 1990](#)).

A SVM considera os conjuntos da forma da Equação 2.2 para alguma constante \mathbf{A} . Na estrutura da Teoria Estatística de Aprendizagem, a constante \mathbf{A} é usada para definir uma estrutura de espaços de hipóteses (o quanto maior for, mais complexo é o espaço da hipótese). O objetivo da SVM é encontrar a solução com a norma EHR “ideal”, ou seja, encontrar o \mathbf{A} ótimo ([EVGENIOU; PONTIL, 1999](#)).

$$\{f : \|f\|_2^k \leq A^2\} \quad (2.2)$$

A segunda opção retratada pelo estudo de [Evgeniou e Pontil \(1999\)](#) é a da função de perda. Uma função de perda é uma função que mapeia um evento ou valores de uma ou mais variáveis num número real, intuitivamente representando algum “custo” associado ao evento ([WALD, 1950](#)). O estudo [Evgeniou e Pontil \(1999\)](#) distingue as SVM em classificação e regressão, porém como este trabalho será focado na classificação, será apenas retratado o primeiro caso, dos classificadores SVM.

Para classificação, o erro de classificação precisa ser minimizado. Então uma função de perda do sinal na forma apresentada pela Equação 2.3, deve ser usada (na classificação \mathbf{y} toma valores binários ± 1 e a classificação é feita tomando o sinal de função $f(x)$). No entanto, devido à escala e às razões computacionais (VAPNIK, 1998), a função de perda real usada para classificação SVM é, Equação 2.4. Ou seja, caso a função de perda seja 0, temos:

$$|1 - yf(x)| < 0$$

caso contrário temos:

$$|1 - yf(x)| > 0$$

Isso também é chamado de função de perda de “margem flexível” por causa de sua interpretação padrão de “margem”: os pontos para os quais a função de perda é zero são os que têm “margem”, descrito pela Equação 2.5 (EVGENIOU; PONTIL, 1999).

$$-yf(x) \tag{2.3}$$

$$|1 - yf(x)|_+ \tag{2.4}$$

$$yf(x) / \|f\|_k^2 \tag{2.5}$$

Para o caso da Equação 2.6 e Equação 2.7. A margem é uma quantidade geométrica importante associada à classificação SVM (EVGENIOU; PONTIL, 1999)

$$1 / \|f\|_k^2 \tag{2.6}$$

$$1 - yf(x) \geq 0 \Rightarrow yf(x) / \|f\|_k^2 \geq 1 / \|f\|_k^2 \tag{2.7}$$

A função para classificação em uma SVM, pode ser generalizada pela Equação 2.8:

$$\min \|f\|_k^2 + C \sum_{i=0}^l |1 - y_i f(x_i)|_+ \tag{2.8}$$

2.11 Random Forest

Desenvolvido por [Breiman \(2001\)](#), o método combina a abordagem de amostragem de Breiman, e a seleção aleatória de características, introduzida independentemente por [Ho \(1995\)](#) e [Amit e Geman \(1997\)](#), a fim de construir uma coleção de árvores de decisão, com variação controlada. Usando ensacamento, cada árvore de decisão no conjunto é construída usando uma amostra com substituição dos dados de treinamento. Estatisticamente, a amostra provavelmente terá cerca de 64% de instâncias aparecendo pelo menos uma vez na amostra. As instâncias na amostra são referidas como instâncias de embalagem e as instâncias restantes (cerca de 36%) são referidas como instâncias fora do pacote. Cada árvore no conjunto funciona como um classificador base para determinar o rótulo de classe de uma instância não rotulada. Isso é feito por meio de votação por maioria, em que cada classificador atribui um voto ao rótulo de classe previsto. Em seguida, o rótulo de classe com o maior número de votos é usado para classificar a instância ([FAWAGREH; GABER; ELYAN, 2014](#)).

O Autor [Pedregosa et al. \(2011\)](#) define *Random Forest* como um meta estimador que se ajusta a vários classificadores de árvore de decisão em várias subamostras do conjunto de dados e usa a média para melhorar a precisão preditiva e controlar o ajuste excessivo. O tamanho da subamostra é sempre o mesmo que o tamanho da amostra de entrada original.

2.11.1 Funcionamento

Segundo [Bastos, Nascimento e Lauretto \(2013\)](#), as *Random Forest* são obtidas através de *bootstrapping aggregating* (ou simplesmente *bagging*), um método utilizado para gerar múltiplas versões de um preditor ([BREIMAN, 1996](#)). Tais versões são construídas a partir de amostras do conjunto original, obtidas via sorteio simples com reposição. Apresentamos a seguir a notação sugerida por [Breiman \(2001\)](#).

Um conjunto de treinamento é denotado por $L = (x_n, y_n), n = 1, 2, \dots, N$, onde N é a quantidade de exemplos, x_n é o vetor de atributos e $y_n \in 1, 2, \dots, C$ é a classe verdadeira do n -ésimo exemplo. Os atributos são indexados por $m = 1, 2, \dots, M$, e assim o vetor de atributos do n -ésimo exemplo é denotado por $x_n = (x_n, 1, x_n, 2, \dots, x_n, M)$. Denote por $\psi(x, L)$ um preditor para a classe de x construído a partir do conjunto de treinamento L .

Suponha que exista uma sequência finita de conjuntos de treinamento $L(k), k = 1, 2, \dots, K$, cada um consistindo de N observações independentes provenientes da mesma distribuição subjacente ao conjunto L . A ideia central é usar $L(k)$ para obter um preditor melhor do que o preditor simples $\psi(x, L)$, tendo como restrição utilizar apenas a sequência de preditores $\psi(x, L(k))$. Indexando-se as classes por $c = 1, 2, \dots, C$, um método de agregar os preditores $\psi(x, L(k))$ e através de votação, escolhendo para x a classe mais votada entre

os preditores.

Formalmente, denotando por $Nc = |\{k \in 1 \dots K : \psi(x, L(k)) = c\}|$ o número de “votos” na classe c , o classificador agregado pode ser definido por $\psi A(x) = \operatorname{argmax}(c) N(c)$. O subscrito A em ψA denota agregação. A obtenção de $L(k)$, $k = 1, 2, \dots, K$ é feita tomando-se amostras *bootstrap* de L , via sorteio com repetição, cada qual de tamanho N . Em cada amostra de treinamento *bootstrap*, aproximadamente 37% das instâncias do conjunto original não são utilizadas para o treinamento (BREIMAN, 1996).

Essas instâncias são usadas como um conjunto de teste, para estimar o erro de cada classificador e, a partir deste, o erro do classificador agregado. O erro *out-of-bag* de cada classificador $\psi(x, L(k))$ é definido como o percentual do conjunto de teste (constituído por $L \setminus L(k)$) classificado erroneamente.

Na formulação das RFs propostas por Breiman (2001), o algoritmo básico de construção das árvores e o CART (sigla do inglês: *Classification and Regression Trees*, Classificação e Árvores de regressão) (BREIMAN, 2017). As árvores são expandidas ao máximo, sem poda. Para a divisão de cada nó, um subconjunto de tamanho fixo dos atributos de entrada é selecionado aleatoriamente, escolhendo-se a divisão ótima dentro desse subconjunto.

2.12 KNN

Segundo Pedregosa et al. (2011), o princípio por trás dos métodos de KNN é encontrar um número pré-definido de amostras de treinamento mais próximas da distância do novo ponto e prever o rótulo a partir delas. O número de amostras pode ser uma constante definida pelo usuário (*k-neighbor neighbor learning*) ou variar com base na densidade local de pontos (aprendizado de vizinho baseado em raio). A distância pode, em geral, ser qualquer medida métrica, a distância euclidiana padrão é a escolha mais comum. Os métodos baseados em vizinhos são conhecidos como métodos de aprendizado de máquina não generalizantes, já que eles simplesmente “lembram” todos os seus dados de treinamento (possivelmente transformados em uma estrutura de indexação rápida, como *Ball Tree* ou *KD Tree*).

Apesar de sua simplicidade, os algoritmos KNN tiveram sucesso em um grande número de problemas de classificação e regressão, incluindo dígitos manuscritos e cenas de imagens de satélite. Sendo um método não paramétrico, muitas vezes é bem-sucedido em situações de classificação em que o limite de decisão é muito irregular (PEDREGOSA et al., 2011).

2.12.1 Funcionamento

Baseando-se no autor [Peterson \(2009\)](#) no qual é apresentada a característica da distância geométrica entre amostras desse algoritmo com uma representação gráfica, a fim de detalhar o funcionamento do mesmo.

O classificador de k mais próximo do vizinho é comumente baseado na distância euclidiana entre uma amostra de teste e as amostras de treinamento especificadas. Utilizando x_{eu} como uma amostra de entrada com p características $(x_{eu1}, x_{eu2}, \dots, x_{eup})$ n ser o número total de amostras de entrada ($i = 1, 2, \dots, n$) e p o número total de recursos ($j = 1, 2, \dots, p$). A distância euclidiana entre a amostra x_i e x_j ($l = 1, 2, \dots, n$) é definido pela Equação 2.9.

$$d(x_i, x_l) = \sqrt{(x_{i1} - x_{l1})^2 + (x_{i2} - x_{l2})^2 + \dots + (x_{ip} - x_{lp})^2} \quad (2.9)$$

Uma representação gráfica do conceito de vizinho mais próximo é ilustrada na tesselação de Voronoi ([VORONOI, 1907](#)) mostrada na Figura. 5. O mosaico mostra 19 amostras marcadas com um "+", e a célula de Voronoi, R , em torno de cada amostra. Uma célula de Voronoi encapsula todos os pontos vizinhos que estão mais próximos de cada amostra.

Uma representação gráfica do conceito de vizinho mais próximo é ilustrada na tesselação de Voronoi ([VORONOI, 1907](#)) mostrada na Figura. 5, na qual apresenta as células de Voronoi de 19 amostras marcadas com um "+". A tesselação de Voronoi reflete duas características do sistema de coordenadas bidimensional do exemplo:

- Todos os pontos possíveis dentro da célula de Voronoi de uma amostra são os pontos vizinhos mais próximos para essa amostra.
- Para qualquer amostra, a amostra mais próxima é determinada pelo Voronoi mais próximo da borda da célula.

O mosaico mostra 19 amostras marcadas com um "+", e a célula de Voronoi, R , em torno de cada amostra. Uma célula de Voronoi encapsula todos os pontos vizinhos que estão mais próximos de cada amostra e é definida pela Equação 2.10.

$$R_i = \{x \in R^p : d(x, x_i) \leq d(x, x_m), \forall i \neq m\} \quad (2.10)$$

Onde R_i é a célula de Voronoi para amostra x_{eu} , e x representa todos os pontos possíveis dentro da célula de Voronoi R_i . Os mosaicos de Voronoi refletem principalmente duas características de um sistema de coordenadas: i) todos os pontos possíveis dentro da célula de Voronoi da amostra são os pontos vizinhos mais próximos para essa amostra e

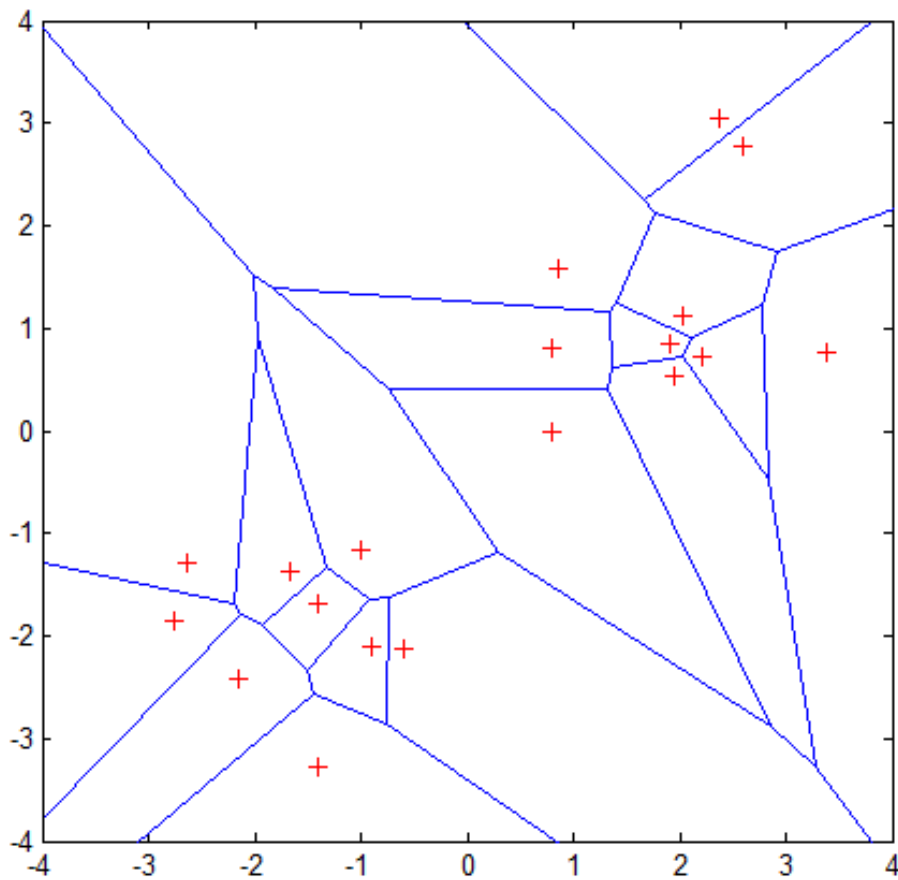


Figura 5 – Tesselação de Voronoi, adaptado de [Voronoi \(1907\)](#).

ii) para qualquer amostra, a amostra mais próxima é determinada pela borda da célula de Voronoi mais próxima. Usando a última característica, a regra de classificação de KNN é atribuir a uma amostra de teste o rótulo da categoria majoritária de suas k amostras de treinamento mais próximas. Na prática, k é geralmente escolhido para ser ímpar, de modo a evitar empates. A regra $k = 1$ é geralmente chamada de regra de classificação de vizinho mais próximo ([PETERSON, 2009](#)).

3 A Doença de Parkinson

3.1 Definição

A DP é uma doença crônica, progressiva e degenerativa do sistema nervoso central, reconhecida principalmente por distúrbios motores como a bradicinesia, os tremores corporais em repouso e a rigidez corporal (COSTA; DIAFÉRIA; BEHLAU, 2016). A bradicinesia é a redução da movimentação, ou seja, o indivíduo consegue se movimentar porém lentamente.

A DP é resultante da degeneração lenta, severa e irreparável dos neurônios dopaminérgicos da substância negra. Os neurônios dopaminérgicos são responsáveis pela produção de dopamina, um neurotransmissor fundamental para a gestão dos movimentos. Com a degeneração destes neurônios, a quantidade de dopamina é reduzida e, consequentemente, as mensagens nervosas para os músculos se tornam gradualmente mais lentas e desorganizadas. A substância negra localiza-se no mesencéfalo, parte intermediária do cérebro, conforme mostra a Figura 6 (EFTAXIAS et al., 2015).

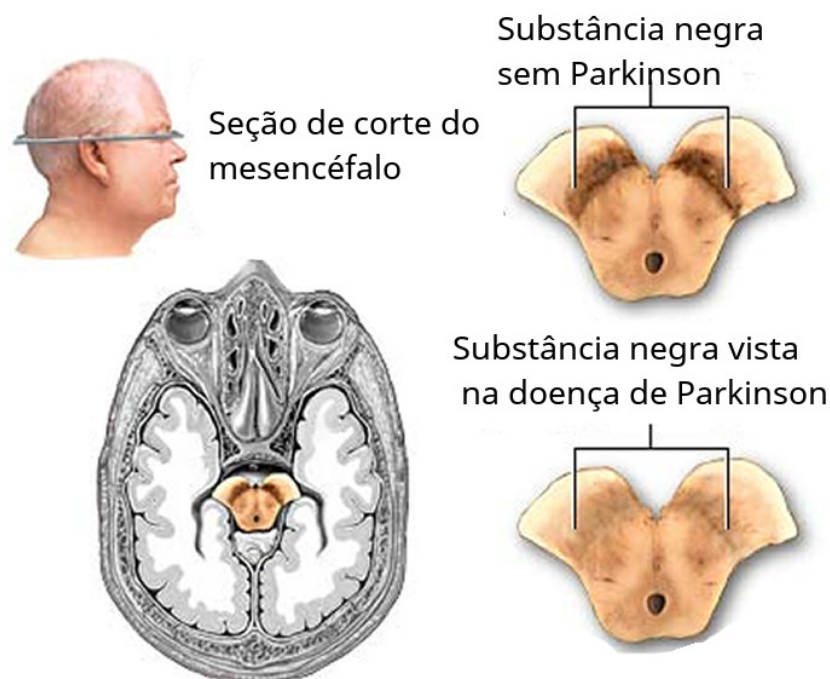


Figura 6 – Diagrama da substância negra, adaptado de Maia (2018).

3.2 Diagnóstico e Sintomas

A DP afeta aproximadamente entre 1% a 2% da população mundial acima dos 65 anos, sendo que no Brasil atinge entorno de 3% da população nesta faixa etária (PETERNELLA; MARCON, 2009). A idade é o único fator de risco conhecido. Segundo Peixinho, Azevedo e Simões (2006) a ocorrência de DP é rara antes dos 40 anos, aumenta após os 50 e tem maior probabilidade após os 70 anos. Porém a incidência da doença não está restrita somente a pessoas idosas, visto que 20% dos casos são de pessoas com menos de 50 anos (GAGO et al., 2014). Não há consenso estabelecido, mas alguns estudos relacionam ser um pouco maior a ocorrência de DP no sexo masculino (PEIXINHO; AZEVEDO; SIMÕES, 2006).

O diagnóstico da DP é clínico, realizado normalmente por um neurologista, onde identifica-se a bradicinesia (lentidão dos movimentos) e pelo menos um de três outros sintomas (GAGO et al., 2014):

- Tremor em repouso;
- Rigidez muscular;
- Instabilidade postural.

Outros exames também podem ser utilizados para confirmação da DP, como tomografia computadorizada e ressonância magnética cerebral, dentre outros a ser prescrito pelo médico responsável (GAGO et al., 2014).

Sobre os sintomas recorrentes da DP, os mesmos são associados a comunicação oral onde 90% dos pacientes possuem problemas relacionados a fala, como gagueira, rouquidão, alteração ou enfraquecimento da voz (ZARZUR et al., 2010). Também pode ocorrer alterações no sono, na memória além de depressão (BARBOSA; SALLEM, 2005). Outro sintoma relevante é a demência que afeta cerca de 25% dos pacientes com DP (PAMPLONA; MATTOS et al., 1996).

3.2.1 Tratamento

A DP não tem cura, consequentemente, os tratamentos disponíveis visam o controle dos sintomas para melhorar a qualidade de vida do paciente (PAMPLONA; MATTOS et al., 1996), com o objetivo de aumentar ao máximo sua autonomia e independência relacionada a atividades do cotidiano. Existem vários tipos de tratamentos, sendo essencial uma equipe multiprofissional, para atendimentos específicos relacionados a cada tipo de sintoma relacionado a doença (SAITO, 2011).

4 Eletromiografia de Superfície

4.1 Definição

A sEMG é o estudo relacionado às transformações elétricas das contrações musculares. É um exame indolor e não invasivo, por permitir a execução com mobilidade, dos movimentos musculares solicitados. A sEMG pode ser executada repetidas vezes sem causar um grande desconforto ao paciente, sendo rápida, livre de radiação e de fácil compreensão. Serve para analisar um grupo ou um feixe muscular específico ([PERNAMBUCO et al., 2010](#)), onde gera-se um sinal sEMG sobre a atividade muscular analisada.

Um sinal sEMG é obtido na medição das tensões relacionados as correntes elétricas, durante a contração muscular. Fornecendo assim em um intervalo de tempo, a média desta atividade neuromuscular ([REAZ; HUSSAIN; MOHD-YASIN, 2006](#)).

A sEMG caracteriza-se pela utilização de um dispositivo sobre a pele do paciente, o qual implica a detecção dos potenciais elétricos relativos às fibras musculares, ou seja, é possível detectar quando um músculo é ativado e qual o movimento foi executado, além de, relacionar a associação dos diferentes músculos envolvidos ([BOTELHO et al., 2010](#)).

4.2 Características do sinal

Para melhor entender as características do sinal gerado pela sEMG, nesta seção é introduzido um breve resumo sobre os elementos que compõe a musculatura humana.

4.2.1 Contração muscular

O processo de ativação muscular começa com a liberação Acetilcolina (ACh - neurotransmissor que opera na comunicação de impulso nervoso do neurônio para as celulares musculares ([SOTO; SEGURA-TORRES, 2005](#)) pelo axônio do neurônio motor até a ligação muscular. Esse neurotransmissor é liberado como consequência do potencial de ação pré-sináptico e resulta na excitação das fibras musculares no PEPS (Processo Pós-Sináptico) devido a abertura dos canais iônicos na membrana plasmática, no caso os receptores nicotínicos, através dos PEPS abertos canais de cálcio necessitantes de tensão elétrica se abrem provocando o potencial de ação pós-sináptico, que se propaga ao longo da fibra muscular ([ROSA, 2005](#)).

A contração muscular ocorre com ligações químicas envolvendo cálcio (Ca^{2+}) e as enzimas troponina, actina e miosina, com a liberação cálcio devido a despolarização nesse processo de propagação. Isto ocorre devido a ligação de cálcio a enzima troponina,

que libera os sítios de ligação na actina para a miosina, provocando o encurtamento das fibras e logo a contração do músculo. A desconexão da miosina com a actina consome ATP (Adenosina Trifosfato), um nucleotídeo encarregado do armazenamento de energia nas ligações químicas (ROSA, 2005).

O relaxamento muscular ocorre quando o cálcio é recapturado pelo retículo sarcoplasmático (retículo endoplasmático especializado no armazenamento de íons de cálcio), por uma bomba de ATP (ROSA, 2005).

4.2.2 Unidade Motora

Primeiramente temos a UM (Unidade Motora), uma coleção de fibras musculares inervadas por um único neurônio motor alfa, sendo a menor unidade operacional de um músculo, podendo ser ativada pelo controle neural. Na atividade muscular, ocorre a ativação do neurônio alfa dentro da UM, produzindo tensão nas fibras musculares ao passo que ocorre a propagação do sinal ao longo das fibras, o músculo é relaxado quando o neurônio alfa para a atividade (YOUSEFI; HAMILTON-WRIGHT, 2014).

O PAUM (Potencial de Ação da Unidade Motora) é a soma dos potenciais de ação das contrações musculares em uma unidade motora relacionadas as tensões detectadas pelos eletrodos da sEMG e descrevem a soma de todas as UM ativas, ou seja, a soma de todos os PAUMs detectadas. Um conjunto de PAUMs que compõem a ação de um músculo é denominado de Trem de PAUM (YOUSEFI; HAMILTON-WRIGHT, 2014). Um diagrama sobre o PAUM pode ser visto na Figura 8.

A Figura 7, representa um esquema simplificado da comunicação do neurônio com a UM do músculo.

A Equação 4.1 descreve um sEMG. Um sEMG é a medida ao longo do tempo da contração de um MUP m , de um total de N , **MUPs**, a função $n(t)$ representa possíveis ruídos encontrados na coleta, ambas as funções são parametrizadas pelo tempo (YOUSEFI; HAMILTON-WRIGHT, 2014).

$$EMG_t = \sum_{m=1}^N MUPs_m(t) + n(t) \quad (4.1)$$

Este tipo de sinal inevitavelmente encontra ruído do equipamento e de outras fontes biológicas presentes no corpo do indivíduo, os quais interferem na coleta (YOUSEFI; HAMILTON-WRIGHT, 2014).

Com relação a DP, o tremor de repouso é o sintoma mais comum, possuindo uma frequência entre 4 e 6 Hz. (JANKOVIC, 2008).

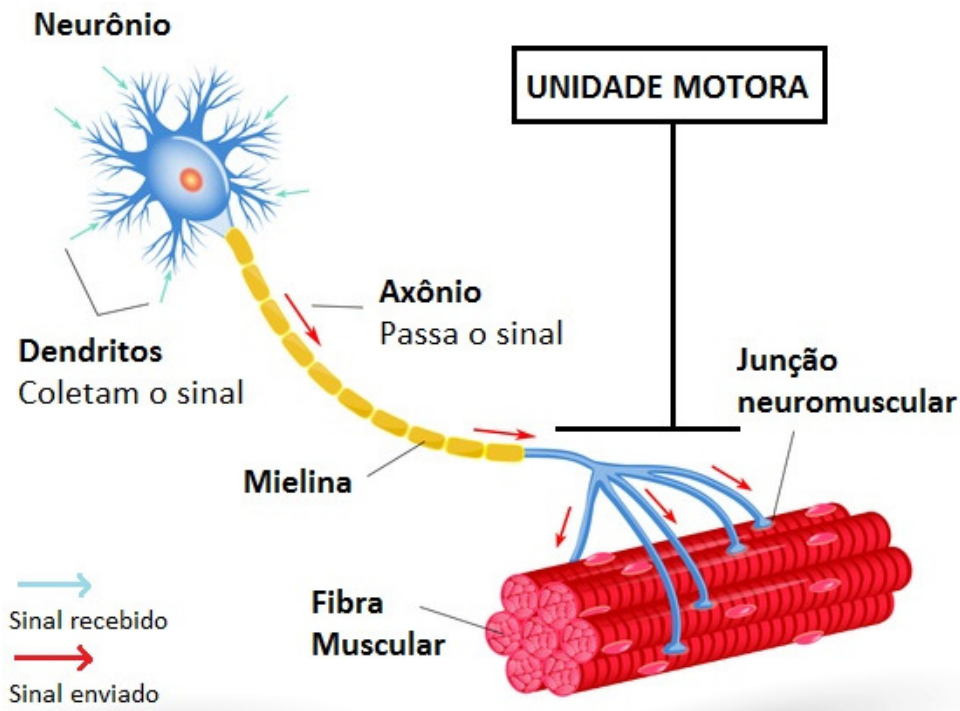


Figura 7 – Modelo de uma unidade motora, adaptado de Sabino (2017).

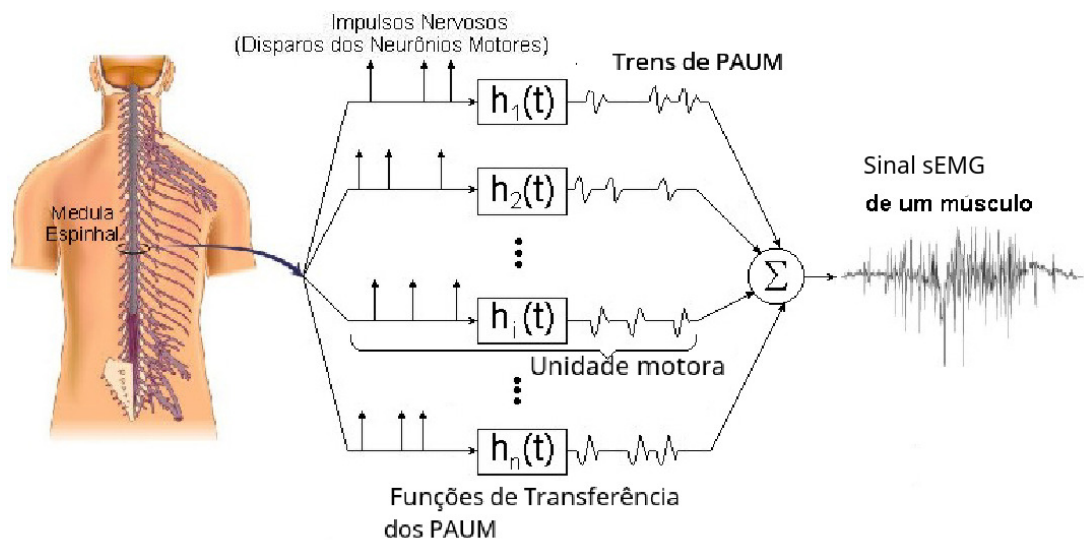


Figura 8 – Sinal sEMG muscular, adaptado de Ortolan (2002).

4.3 Aplicações da sEMG

A sEMG é largamente utilizada por profissionais de diversas áreas científicas que estudam o movimento humano, como médicos, fisioterapeutas, fonoaudiólogos e profissionais em educação física, sendo utilizada na análise clínica relacionada a ativação muscular, intensidade e duração durante a contração (NASCIMENTO et al., 2012). Tais informações podem ser utilizadas, por exemplo, no estudo da fadiga muscular em atletas (SCHMITZ et al., 2012) ou na detecção de doenças degenerativas como DP (EFTAXIAS et al., 2015).

5 Trabalhos Relacionados

5.1 AM, DP e sEMG

Como observado por [Yousefi e Hamilton-Wright \(2014\)](#), os estudos relacionados a sEMG e AM são separados em dois tipos de transtornos neuromusculares, miopatia e neuropatia. Onde miopatia refere-se a um grupo de patologias que atingem diretamente o tecido muscular sem relação com disfunções do sistema nervoso. Já a neuropatia, onde encontra-se a DP, é caracterizada por qualquer dano nos nervos envolvidos no controle muscular. Segundo o autor [Yousefi e Hamilton-Wright \(2014\)](#) as diversas técnicas de AM sobre sEMG, pode-se catalogá-las em três etapas, análise, decomposição e classificação, sendo que não é necessariamente sequencial, podendo ser um processo iterativo.

Outro estudo com bons resultados, porém sem relação direta com a DP, foi desenvolvido por [Katsis et al. \(2006\)](#), que propôs um método para detectar automaticamente um PAUM e classificá-lo em normal, neuropático ou miopático. O método possui três etapas, pré-processamento do sinal sEMG, detecção e agrupamento do PAUM, e posteriormente a classificação deste utilizando uma SVM. Como resultados foi possível identificar corretamente para o agrupamento 93%, 95% e 92% respectivamente, e para a classificação do PAUM foi alcançado aproximadamente 86%.

5.1.1 Análise

O pré-processamento ou análise dos dados é uma das partes mais importantes do processo de AM. Essa etapa ocorre após a coleta dos dados, possuindo alguns objetivos como organizar os dados em conjuntos, identificar possíveis problemas como ruídos ou valores desconhecidos. Também é comum utilizar esta etapa para simplesmente aprender mais sobre os modelos, plotando gráficos e colhendo mais informações. Também utiliza-se essa etapa para realizar tratamento nos dados, como transformações lineares para um conjunto de dados mais simples de ser trabalhado ([BATISTA et al., 2003](#)).

5.1.2 Decomposição

Segundo [Yousefi e Hamilton-Wright \(2014\)](#) a decomposição de um sinal sEMG consiste de cinco etapas, aquisição, segmentação, extração de *features*, agrupamento de MUPs e atribuição de MUP.

A seleção das *features* ou características é a distinção de um sinal bruto em informações úteis, esta informação tratada consiste uma *feature*, a qual será utilizada para

treinar o modelo, além disto busca-se remover a parte indesejada do sinal e interferências associadas (PHINYOMARK; PHUKPATTARANONT; LIMSAKUL, 2012).

Segundo Phinyomark, Phukpattaranont e Limsakul (2012), essa etapa deve ser trabalhada cuidadosamente, uma vez que erros podem prejudicar a classificação dos dados. Em sua pesquisa sobre as *features* em sEMG, verificou-se utilizando gráficos de dispersão, análise estatística e classificação, que a maioria das *features* no domínio do tempo são redundantes e desnecessárias, sendo melhor agrupadas em quatro tipos: energia e complexidade, frequência, modelo de previsão e dependência do tempo. O estudo analisou 37 métodos de extração de *features* e recomendou o uso das seguintes *features*:

- **MAV**: Sigla do inglês: *Mean absolute value*, Valor absoluto médio. É uma média do valor absoluto da amplitude de um sinal EMG em um segmento, é uma das *features* mais utilizadas em análise de sinais EMG. Utilizado como método de informação de energia.
- **WL**: Sigla do inglês: *Waveform length*, Comprimento da forma da onda. É definido como o comprimento da forma de onda EMG, ao longo do tempo, também é conhecido como comprimento de onda. Utilizado como método de informação de complexidade;
- **WAMP**: Sigla do inglês: *Wilson amplitude*, Amplitude de Willison. É definido como a diferença entre a amplitude do sinal EMG entre dois segmentos adjacentes que excede um limite pré-definido. Utilizado como método de informação de frequência;
- **AR**: Sigla do inglês: *Auto-regressive coefficients*, Amplitude de Willison. É um modelo de predição que descreve cada amostra do sinal EMG como uma combinação linear das amostras anteriores. Utilizado como método de modelo de predição;

Outro estudo realizado por Yousefi e Hamilton-Wright (2014), cita algumas *features* compiladas de outros trabalhos, como observado na Figura 9, sendo elas:

- **Amplitude pico a pico**: relaciona os valores nas extremidades mínima e máxima de ativação de uma MUP;
- **Tempo de ativação**: descreve a duração entre início e término de ativação de uma MUP;
- **Área de uma MUP**: é obtida integrando a área do sinal gerado, resultante da duração do movimento de uma MUP;
- **Espessura**: é a razão entre a área e a amplitude.

No trabalho realizado por [Yousefi e Hamilton-Wright \(2014\)](#), um bom resultado foi obtido na decomposição, onde realizou-se uma classificação utilizando o algoritmo SVM, o qual obteve uma acurácia de 70,4% no modelo, utilizando uma técnica de redução de dimensionalidade para a extração das *feature*, a DWT (sigla do inglês: *discrete wavelet transform*, transformada discreta de *Wavelets*). Este método utiliza uma função capaz de remodelar uma função qualquer no domínio do tempo, em diferentes escalas de frequência e tempo, após a utilização desta técnica o algoritmo SVM obteve uma acurácia de 99,3%.

Com relação a redução de dimensionalidade, no estudo desenvolvido por [Güler e Koçer \(2005\)](#), realizou-se um processo de classificação utilizando SVM e sEMG e para a redução de dimensionalidade utilizou-se a combinação entre a FFT (sigla do inglês: *fast Fourier transform*, Transformada rápida de Fourier), combinada com a PCA (Análise de Componentes Principais, sigla do inglês: *Principal Component Analysis*). Esta combinação é utilizada para facilitar o cálculo e armazenamento dos dados do sinal sEMG. Por fim, em [Güler e Koçer \(2005\)](#) foi concluído que a SVM possui um alto nível de predição com relação aos sinais sEMG.

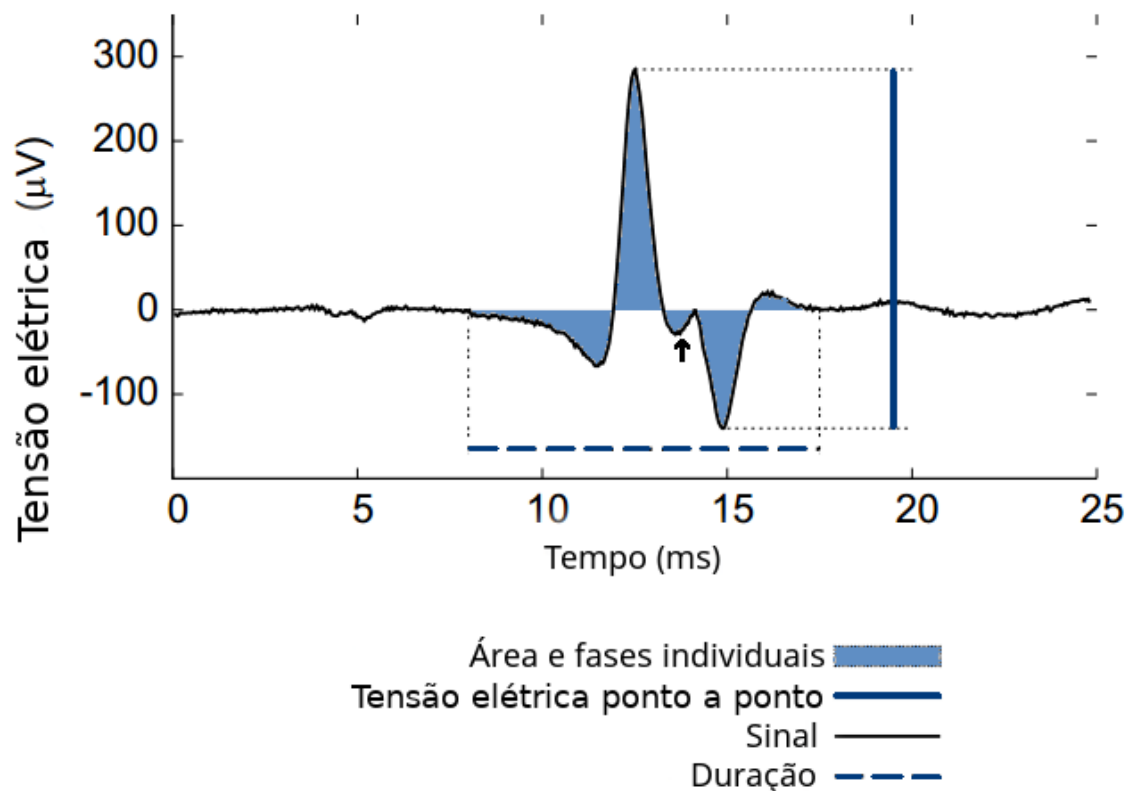


Figura 9 – *Features* de um sEMG, adaptado de [Yousefi e Hamilton-Wright \(2014\)](#).

Outro trabalho com foco no estudo relacionado a eficiência dos métodos de redução de dimensionalidade foi realizado por [Camara et al. \(2015\)](#) que utilizou uma rede neural artificial para identificar a doença através do tremor em repouso.

No estudo realizado por [Ai, Wang e Yao \(2011\)](#), utilizou-se sEMG com SVM para identificar a DP e o tremor essencial. Ele utilizou como técnica de redução de dimensionalidade, uma EMD (sigla do inglês: *empirical mode decomposition*, decomposição em modo empírico) para decompor o sinal em IMF (sigla do inglês: *intrinsic mode functions*, funções do modo intrínseco). E seguindo, decompôs as IMF em SVD (sigla do inglês: *singular value decomposition*, decomposição em valores singulares) para extrair as *features* das IMF geradas, com este tratamento os dados são melhores calculados e processados pelo algoritmo SVM. Outro método de redução de dimensionalidade foi aplicado por [Ai, Wang e Yao \(2011\)](#), no qual foi comparado o desempenho comutando a EMD pela DWT e verificou através de validação cruzada que o método EMD-SVD foi superior ao método DWT-SVD.

A SVM também foi utilizada por [Kugler et al. \(2013\)](#), que combinou sEMG com sinais de um acelerômetro, para identificar a DP e o Tremor essencial, neste trabalho utilizou-se a validação cruzada, para verificar a integridade dos resultados.

Outro trabalho interessante foi realizado por [Loconsole et al. \(2018\)](#), que propôs uma técnica em 3 estágios para diferenciar portadores da DP, foi utilizada uma análise caligráfica com sEMG, para a classificação dos dados, foi utilizado uma SVM alinhada com uma RNA.

Com relação as *features*, um comparativo sobre as estratégias de redução de dimensionalidade em sEMG foi realizado por [Liu \(2014\)](#), obtendo uma média de 95% de precisão na classificação em com 12 *features* selecionadas.

6 Solução de software

Esse projeto utilizará a metodologia ágil de desenvolvimento de *software* com uma adaptação do *Scrum* para projetos pequenos, para a organização e desenvolvimento do trabalho como um todo.

6.1 Ágil

6.1.1 *Scrum*

O *Scrum* é uma metodologia ágil para o gerenciamento de projeto de *software*, que consiste na estruturação de um projeto em pequenas fases incrementais, com prazo definido e escopo variável de acordo as demandas e necessidades do cliente. As iterações são chamados de *Sprints* e normalmente vão de uma semana a um mês, dependendo da organização e necessidades do projeto. As funcionalidades do *software* a serem desenvolvidas, são colocadas em uma lista denominada *Product Backlog*. Em cada *Sprint* as funcionalidade são priorizadas e selecionadas, e após são alocadas na lista de *Sprint Backlog*, as quais devem ser desenvolvidas na *Sprint* (SUTHERLAND, 2016).

Durante as *Sprints*, devem ocorrer reuniões diárias, as quais necessitam ser objetivas e rápidas, com tempo fechado, não sendo ideal ultrapassar 15 minutos. Nessa reunião, busca-se alinhar o desenvolvimento do projeto entre toda a equipe, onde todos os integrantes devem responder as seguintes questões (SUTHERLAND, 2016).:

- *O que eu fiz ontem?*
- *O que eu farei hoje?*

Ao final de cada *Sprint*, devem ocorrer uma reunião de retrospectiva, onde a equipe planejará a nova *Sprint* (SUTHERLAND, 2016).

6.2 Ferramentas

Como descrito na seção 2.7 na página 31, uma das principais linguagens de programação em AM é o *Python*, com a utilização da biblioteca *scikit-learn*. Com este cenário em vista, optou-se por desenvolver o resto da aplicação em *Python*. Deste modo, deve-se obter uma boa integração com o modelo desenvolvido.

6.2.1 Scikit-learn

O *scikit-learn* será o *framework* utilizado para o desenvolvimento dos modelos de AM.

6.2.2 Django-Rest

A *Django Rest Framework* é uma biblioteca em *Python* que viabiliza de forma simples a criação de APIs com o padrão arquitetural REST (*sigla do inglês: Representational State Transfer, Transferência de Estado Representacional*) em projetos Django usando padrões estabelecidos para proporcionar produtividade na criação de APIs. Sendo que, REST é um padrão arquitetural da engenharia de *software* utilizado para a comunicação entre diferentes sistemas, e Django é um *framework* para a criação de páginas *web* (CHRISTIE, 2011).

6.2.3 React

A *React* é uma biblioteca *JavaScript* declarativa criada pelo *Facebook*. Sendo bem eficiente e flexível para a criação de interfaces de usuário. Ou seja, ela é simplesmente uma coleção de funcionalidades que podem ser chamadas pelo desenvolvedor, para desenvolver uma interface de usuário reaproveitável (Facebook Inc, 2018).

6.3 Produto Mínimo Viável

Um MVP (*sigla do inglês: minimum viable product, produto viável mínimo*) é uma técnica de desenvolvimento na qual um novo produto ou site é desenvolvido com recursos suficientes para satisfazer os primeiros usuários. O conjunto final e completo de recursos é projetado e desenvolvido somente após considerar o feedback dos usuários iniciais do produto (TECHOPEDIA, 2018).

A solução de *software* presente neste trabalho de conclusão de curso é um MVP que consiste no núcleo principal de um possível produto final a ser desenvolvido. São especificados no Documento de Visão, documento apresentado no Apêndice B, os requisitos que atendem as principais necessidades dos usuários vide ao que foi proposto neste trabalho. Além disso o Apêndice B apresenta os perfis do próprio usuário no qual o MVP do produto será destinado e das partes interessadas no projeto como um todo.

7 Metodologia

7.1 Comitê de Ética e Pesquisa

A coleta de dados foi autorizada pelo Comitê de Ética e Pesquisa da Faculdade da Saúde da UnB CAAE 38386714.8.0000.0030 e foi realizado pelos pesquisadores Jorge Luiz (Engenheiro Eletricista) e Bruna da Silva (Fisioterapeuta).

Ela utiliza o protocolo descrito no Anexo A, o qual descreve detalhadamente todo o processo de coleta dos dados, bem como o Termo de Consentimento Livre e Esclarecido – TCLE encontra-se no Anexo B.

7.2 Processo de construção de uma aplicação de Aprendizado de Máquina

Segundo Amazon (2018) e Géron (2017), o processo de desenvolvimento de uma aplicação de AM é iterativo e envolve uma série sequencial de passos, que de maneira geral vão desde entender o quadro geral até a apresentação da solução da proposta da aplicação. O processo pode ser visualizado pela representação BPMN (sigla do inglês: *Business Process Model and Notation*, Modelo e Notação de Processos de Negócio) na Figura 10.

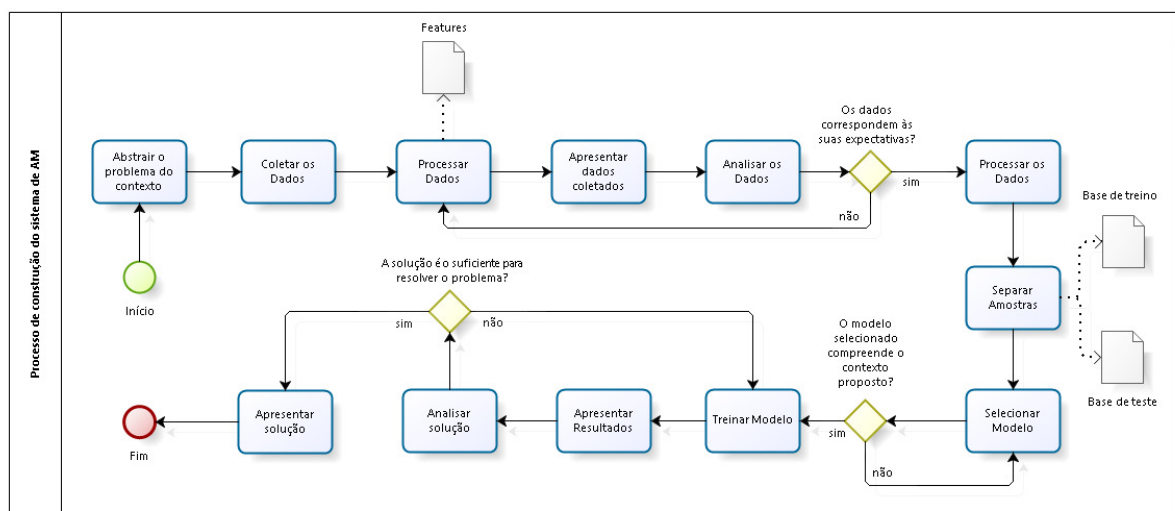


Figura 10 – Diagrama do processo (Imagem dos autores).

7.2.1 Abstrair o problema no contexto

O primeiro passo é entender o que será predito, ou seja, analisar a natureza do problema observando quais respostas o modelo escolhido deverá prever, para auxiliar na solução do problema (GÉRON, 2017).

Tendo em vista o seguinte cenário, no qual é necessária a fabricação de produtos, porém a decisão de qual produto fabricar depende do número de vendas em potencial. Nesse cenário, é necessário saber a quantidade de vezes que cada produto foi comprado, ou seja, prever o número de vendas. Existem diversas maneiras de definir esse problema utilizando AM, porém a escolha de como definir o problema depende do caso de uso ou necessidade comercial (AMAZON, 2018).

Um ponto importante do entendimento do contexto é a abstração do problema. Essa etapa tem um papel fundamental na escolha dos modelos, pois tendo se entendido a natureza do problema, será selecionado o modelo para prever os resultados com base no tipo de aprendizado mais adequado a este contexto. Além disso, é necessário definir claramente quais os objetivos deverão ser atingidos, qual a métrica de desempenho utilizar para validar o modelo e quais algoritmos utilizar para alcançar o desempenho esperado. Com estas e outras respostas será possível enquadrar o problema em supervisionado, não supervisionado ou por reforço e após isso categoriza-lo em classificação, regressão (GÉRON, 2017). Essa etapa de análise tem como finalidade orientar o trabalho evitando a criação de modelos que não respondem o problema, além de otimizar o esforço no desenvolvimento (AMAZON, 2018).

7.2.2 Coleta de dados

Os dados coletados nessa etapa devem estar ligados diretamente ao que é requerido pelo problema. Atender esses requisitos a fim de atender as necessidades do contexto é de suma importância, pois alguns dos já citados desafios encontrados no desenvolvimento de sistema de AM estão relacionados aos dados que serão utilizados (GÉRON, 2017).

A quantidade de dados coletados nessa etapa ainda é incerta, porém é levado em consideração que quanto maior a quantidade de dados coletados, melhor. Neste trabalho são utilizados sinais captados por dispositivos sEMG, mais informações sobre esse sinal são apresentadas no capítulo 4.

7.2.3 Analisar os Dados

Antes de passar os dados para o algoritmo de AM, é considerado uma boa prática inspecionar seus dados para identificar problemas e obter *insights* sobre os dados que se está utilizando (AMAZON, 2018), quanto melhor forem os dados, melhor será o modelo preditivo que está utilizando esses dados.

Segundo [Amazon \(2018\)](#), enquanto você está analisando seus dados é importante responder às seguintes perguntas:

- Os dados correspondem às suas expectativas?
- Houveram problemas nas coletas dos dados?
- Quais classes são as mais frequentes dentre os seus dados?
- Existem mais valores ausentes ou inválidos do que o esperado?

Além de construir um resumo dos dados relacionados a essas perguntas, também é uma boa prática conhecer a correlação entre cada variável e a classe de destino. Em geral, é necessário incluir variáveis com alta correlação, porque elas possuem maior poder preditivo (sinal) e deixam de fora as variáveis com baixa correlação, porque são provavelmente irrelevantes ([AMAZON, 2018](#)).

Essas práticas são importantes para conhecer os dados que serão utilizados podendo prevenir ou facilitar a identificação de possíveis erros futuros com relação a utilização dos dados.

7.2.4 Processar Dados

Após a etapa de análise dos dados, é necessário tornar as variáveis mais significativas. Esta etapa é conhecida como processamento de *features*, sendo responsável por preparar os dados para serem consumidos pelos algoritmos de AM e pode ser resumida em três etapas mais comuns ([BLUM; LANGLEY, 1997](#)):

1. **Formatar:** Os dados selecionados podem não estar no formato desejado ou os dados estão em um banco de dados relacional e é desejado que os mesmos estejam em um arquivo simples. Formatar os dados para um padrão que facilita o trabalho com esses dados é a primeira etapa do processamento.
2. **Limpar:** Remover ou consertar os dados que faltam dentro de um banco de dados. Pode haver instâncias de dados incompletas e não conter os dados para solucionar o problema. Essas instâncias podem precisar ser removidas. Além disso, pode haver informações confidenciais em alguns dos atributos e esses atributos podem precisar ser anonimizados ou removidos dos dados completamente.
3. **Selecionar Amostras:** No caso de se trabalhar com uma grande massa de dados, nem sempre todos esses dados serão utilizados. A quantidade exagerada de dados pode levar a tempos de execução muito elevados e maiores requisitos computacionais. Selecionar uma amostra representativa dos dados selecionados pode ser muito mais rápida para explorar os resultados.

7.2.5 Selecionar o Modelo

Após identificar o problema, explorar os dados, limpar as incoerências e separar os dados em treino e teste, é possível escolher o modelo adequado para solucionar o problema identificado (GÉRON, 2017).

7.2.6 Treinar o Modelo

Para treinar o modelo, com a totalidade do dados separados em, por exemplo, 70% dados de treino e 30% em dados de teste. Utiliza-se os 70% dos dados para treinar o modelo, passando estes dados pelo algoritmo escolhido, e os outros 30% restantes serão utilizados para testar a eficiência do modelo.

7.2.7 Validar a solução

Para validar a solução, utiliza-se a validação cruzada, a qual foi descrita na seção 2.9. Também é comum a utilização da matriz de confusão descrita na seção 2.8.1.

7.2.8 Apresentar a solução

A tecnologia utilizada para o desenvolvimento neste trabalho, *Python*, possui muitas bibliotecas complementares para fazer visualizações estáticas ou dinâmicas, mas focaremos principalmente na combinação entre *Matplotlib*, *Numpy* e *Pandas*.

8 Resultados

8.1 Processo de AM

Esta seção descreve a realização do processo de AM, deste a coleta até a apresentação dos dados.

8.1.1 Coleta

A coleta de dados foi autorizada pelo Comitê de Ética e Pesquisa da Faculdade da Saúde da UnB CAAE 38386714.8.0000.0030. Os dados utilizados para a classificação, são referentes a **Coleta 1 - Tremor de Repouso**, uma vez que esta apresentou os melhores resultados na classificação, já que as outras três coletas apresentaram um alto ruído que acabaram comprometendo os dados. A seguir um pequeno resumo sobre esta coleta.

Foram realizadas três coletas por colaborador, com duração de cinco segundos cada, a disposição dos eletrodos encontra-se no Anexo B. Os dados do sinal foram armazenadas através da utilização de 4 canais, sendo eles:

1. **CH1**: Extensor radial do longo do carpo direito;
2. **CH2**: Flexor superficial dos dedos direito;
3. **CH3**: Extensor radial do longo do carpo esquerdo;
4. **CH4**: Flexor superficial dos dedos esquerdo.

Estes dados coletados em sEMG em cada canal, foram cedidos no formato *edf*. Sendo que, os grupos consistiam em 14 pessoas no grupo de controle, cada um com três amostras e 14 pessoas no grupo de pessoas com a DP, também com três amostras cada.

8.1.2 Processamento dos dados

Com os dados brutos foi iniciada a etapa de pré-processamento dos dados. Inicialmente padronizou-se os dados utilizando os pontos entre 2000 a 11000, pegando assim somente os dados úteis, uma vez que os dados não tinham a mesma continuidade. Utilizou-se como *feature* os sinais obtidos pela FFT, onde se realizou testes nos quatro canais disponíveis, realizando um comparativo da acurácia com e sem os dados obtidos a partir da FFT, além de um comparativo sobre a acurácia entre os canais, um exemplo do sinal pode ser observado na Figura 11, já o sinal bruto pode ser observado na Figura 12.

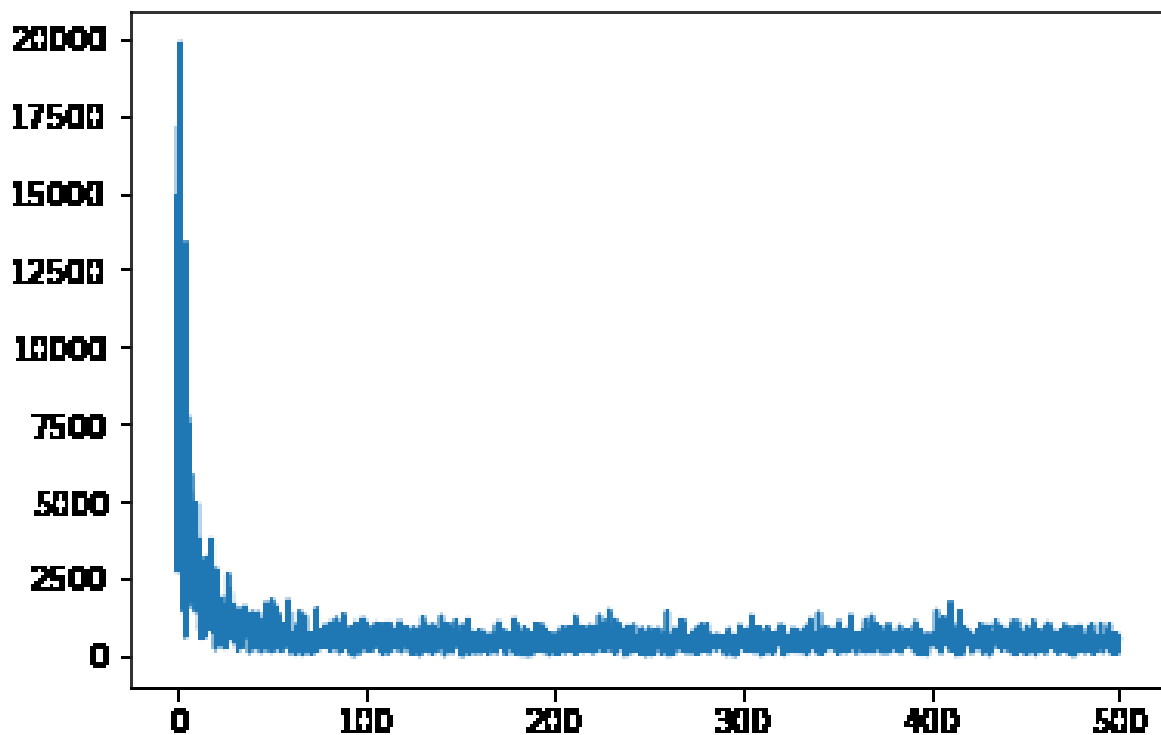


Figura 11 – Amostra de um sinal após a aplicação do FFT no canal 1 (Imagem dos autores).

Tabela 1 – Comparativo da acurácia entre os canais com a FFT.

	Parkinson	Controle	Em Conjunto
CH1	86%	67%	78%
CH2	64%	50%	58%
CH3	71%	67%	69%
CH4	71%	25%	59%

Como a FFT retorna números complexos ($a + bi$), utilizou-se o módulo para cada um dos valores.

Com relação aos dados obtidos em cada canal, como pode ser observado na Figura 1. Utilizando a FFT obteve-se um bom resultado nos canais **CH1** com uma precisão média de 78% e um resultado razoável no **CH3** com precisão média de 69%. Já o **CH4** obteve uma precisão insatisfatória de 59% de acerto, resultado semelhante ao **CH2** que teve a precisão de 58%.

Uma Visão mais detalhada sobre este resultado pode ser observada nas matrizes de confusão na Figura 13 e na Figura 14. Com relação ao **CH1** obteve-se uma precisão de 86% com 14% de falsos positivos no grupo de DP e uma precisão de 67% de acerto e 33% de falsos positivos no grupo de controle. Já o canal **CH2** obteve um péssimo resultado no grupo de controle com somente 50% de acerto.

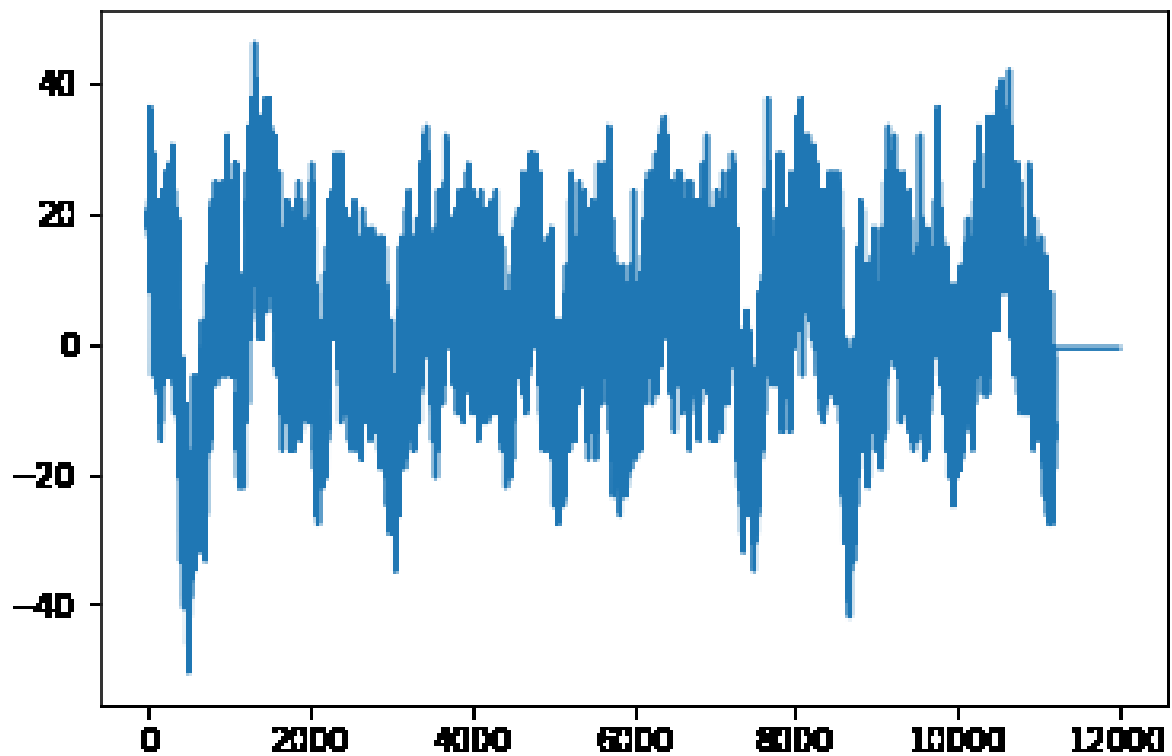


Figura 12 – Amostra de um sinal bruto no canal 1 (Imagem dos autores).

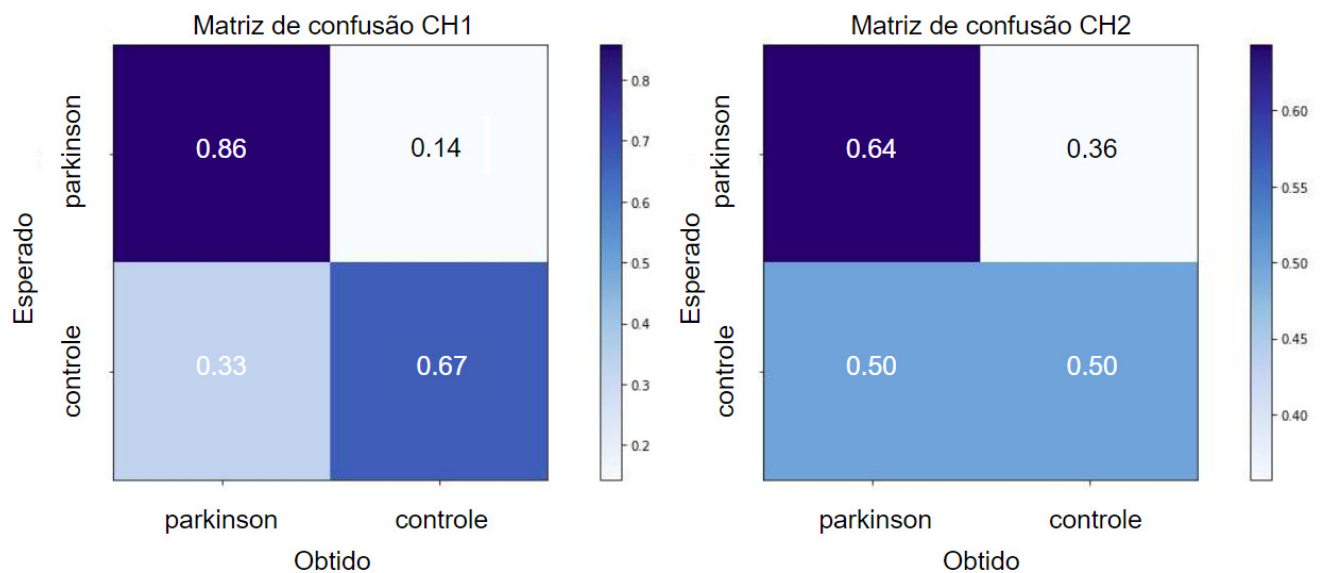


Figura 13 – Comparativo entre os canais 1 e 2, com a FFT (Imagem dos autores).

Outro canal que teve um resultado ruim foi o canal **CH4**, mesmo com 71% de acerto no grupo com a DP, obteve 75% de falsos positivos no grupo de controle. Já o **CH3** obteve um resultado razoável com precisão de 71% no grupo com DP e e 67% no grupo de controle.

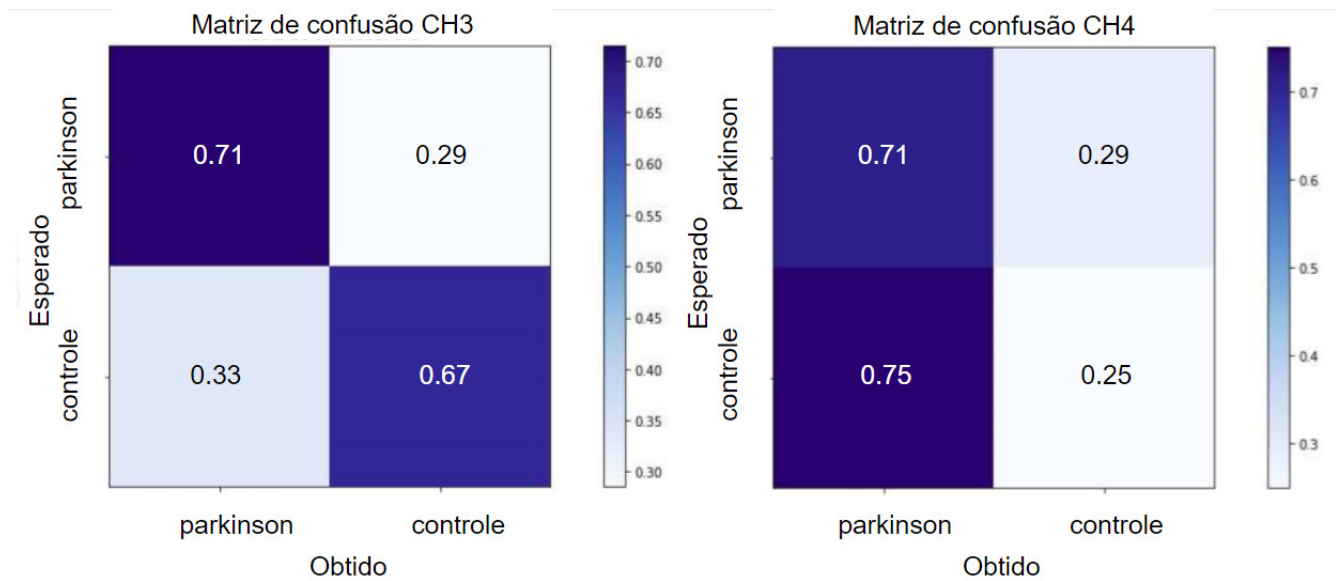


Figura 14 – Comparativo entre os canais 3 e 4, com a FFT (Imagem dos autores).

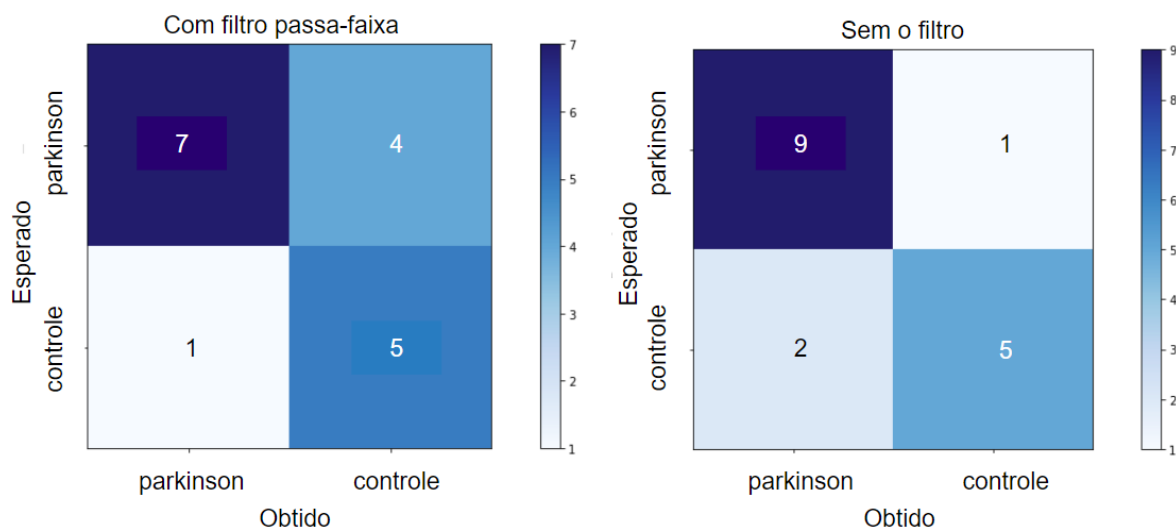


Figura 15 – Matriz de confusão com o filtro passa-banda e sem o filtro no **Ch1** (Imagem dos autores).

Realizou-se também a aplicação de filtros passa-banda, passa-baixa e passa-alta. Porém estes filtros diminuíram a acurácia obtida, como observado na Figura 15, que exemplifica a matriz de confusão referente ao **CH1**, assim como normalizou-se os dados em diferentes escalas e do mesmo modo a acurácia caiu.

Na Figura 15 tem-se que, com a utilização do filtro passa-banda com um total de 11 amostras de pessoas com DP, o sistema conseguiu classificar corretamente 7 amostras ou 64%, e no grupo de controle classificou-se corretamente 5 amostras de um total de 6 ou 83%. Com relação aos testes sem a utilização do filtro, obteve-se uma previsão de 90% no grupo de DP e 71% no grupo de controle, portanto superior aos testes utilizando o filtro.

Tabela 2 – Comparativo entre a acurácia dos algoritmos do **Ch1**.

	SVM	Random Forest	KNN
Somente FFT	80,77%	69,23%	53,85%
FFT com PCA	84,62%	80,77%	53,85%
Somente FFT com <i>Cross-validation</i>	68,00%	75,00%	59,00%
FFT com PCA e <i>textitCross-validation</i>	74,00%	79,00%	59,00%

Também realizou-se a classificação utilizando todos os canais ao mesmo tempo, porém o resultado foi insatisfatório.

Outra técnica de seleção de *feature* utilizada foi a PCA, com a utilização desta técnica obteve-se uma leve melhora na acurácia, como pode ser observado na Tabela 2, a qual utilizou-se o *cross-validator* para verificar a precisão real do modelo.

Como pode ser observado na Tabela 2, com o auxílio do k-fold, pode-se inferir que o *Random Forest* possui uma precisão superior ao SVM, além de que o PCA aumenta a acurácia consideravelmente do modelo.

Finalizando a etapa de pré-processamento dos dados, optou-se pela utilização da FFT com o PCA, uma vez que, esta combinação obteve os melhores resultados. Utilizou-se então o *Random Forest* devido a sua acurácia ser levemente superior ao SVM e o consideravelmente superior ao KNN.

8.2 Solução de *software*

O *software* pode ser acessado livremente no GitHub, através do seguinte repositório <https://github.com/SkiNgK/mlModels>, na pasta **sistema**, divididos em *backend* e *frontend*. Os detalhes referentes a implementação estão descritos nas seções seguintes.

8.2.1 Arquitetura

O documento de Arquitetura encontra-se no Apêndice A.

8.2.2 Visão

O documento de Visão encontra-se no Apêndice B.

8.2.3 Produto

Os detalhes referentes a construção do produto encontram-se descritos no documento de Visão (Apêndice B) e no documento de Arquitetura (Apêndice A).

8.2.3.1 Pagina inicial

A Figura 16 exemplifica a página inicial da aplicação, sendo que esta aplicação conterá somente três páginas. Além da página inicial tem-se a página de pacientes arquivados e a página sobre com as informações relacionadas a metodologia e resultados sobre o processo de Aprendizado de Máquinas utilizado no desenvolvimento da aplicação.

<input type="checkbox"/>	Id	Nome	Idade	Sexo	Resultado sobre a DP	Opções
<input type="checkbox"/>	62	Maria ferandez	23	Masculino	AVALIAR	EDITAR REAVALIAR
<input type="checkbox"/>	61	José raimundo alvares fernandes	23	Masculino	AVALIAR	EDITAR REAVALIAR
<input type="checkbox"/>	60	velho novo com parkinsom	23	Masculino	AVALIAR	EDITAR REAVALIAR
<input type="checkbox"/>	59	velho novo com parkinsom	23	Masculino	Baixa probabilidade	EDITAR REAVALIAR
<input type="checkbox"/>	58	teste de velho paciente	33	Masculino	Baixa probabilidade	EDITAR REAVALIAR
<input type="checkbox"/>	57	teste de velho paciente	33	Masculino	Alta probabilidade	EDITAR REAVALIAR
<input type="checkbox"/>	56	novo paciente com parkinsom	23	Masculino	Alta probabilidade	EDITAR REAVALIAR
<input type="checkbox"/>	55	novo paciente	22	Masculino	Baixa probabilidade	EDITAR REAVALIAR
<input type="checkbox"/>	54	bancada	46	Masculino	Alta probabilidade	EDITAR REAVALIAR
<input type="checkbox"/>	50	Pedro Alvares Cabral	15	Masculino	Baixa probabilidade	EDITAR REAVALIAR

Linhas por página: 10 1-10 de 13

Informações
 Software, capaz de auxiliar o diagnóstico da DP (Doença de Parkinson) avaliando tremores patológicos de repouso, utilizando AM (aprendizado de máquinas) para identificar se o indivíduo é um possível portador ou não da doença, utilizando dados coletados através da sEMG (sigla do inglês: surface electromyography, eletromiografia de superfície).

Figura 16 – Pagina inicial (Imagem dos autores).

8.2.3.2 Cadastro de pacientes

A Figura 17 refere-se ao cadastro de pacientes. Nesta tela inserem-se os seguintes itens, todos obrigatórios:

- Nome completo;
- Idade;
- Sexo;
- Arquivo sEMG em formato *edf*.

HOME ARQUIVADOS SOBRE Pesquisa por nome Cadastro de paciente Sign Up

Cadastro de paciente

Nome
nome completo

Idade
idade

Sexo
Masculino

Insira o arquivo EMG aqui!
Escolher arquivo Nenhum arquivo selecionado

Cancelar Salvar

Lista de pacientes

<input type="checkbox"/>	Id	Nome
<input type="checkbox"/>	62	Maria ferandez
<input type="checkbox"/>	61	José raimundo alvares f
<input type="checkbox"/>	60	velho novo com parkins
<input type="checkbox"/>	59	velho novo com parkins
<input type="checkbox"/>	58	teste de velho paciente
<input type="checkbox"/>	57	teste de velho paciente
<input type="checkbox"/>	56	novo paciente com park
<input type="checkbox"/>	55	novo paciente
<input type="checkbox"/>	54	bancada
<input type="checkbox"/>	50	Pedro Alvares Cabral

Linhas por página: 10 1-10 de 13

Informações
Software, capaz de auxiliar o diagnóstico da DP (Doença de Parkinson) avaliando tremores patológicos de repouso, utilizando AM (aprendizado de máquinas) para identificar se o indivíduo é um possível portador ou não da doença, utilizando dados coletados através da sEMG (sigla do inglês: surface electromyography, eletromiografia de superfície).

Figura 17 – Cadastro das pacientes.

8.2.3.3 Edição de pacientes

Para realizar a edição dos dados de um paciente, como demonstrado na Figura 17, ao pressionar o botão **EDITAR** na aba de opções é aberto um *modal* contendo as informações do paciente para a alteração. Para alterar o arquivo sEMG é necessário clicar em alterar arquivo, o qual após adicionado substitui o antigo.

8.2.3.4 Avaliar pacientes

Para avaliar um paciente recém inserido é necessário clicar no botão **AVALIAR**, como na Figura 19.

8.2.3.5 Pesquisar pacientes

Para pesquisar por qualquer um dos itens das colunas, basta clicar no ícone de pesquisa como na Figura 20 e inserir o termo da pesquisa.

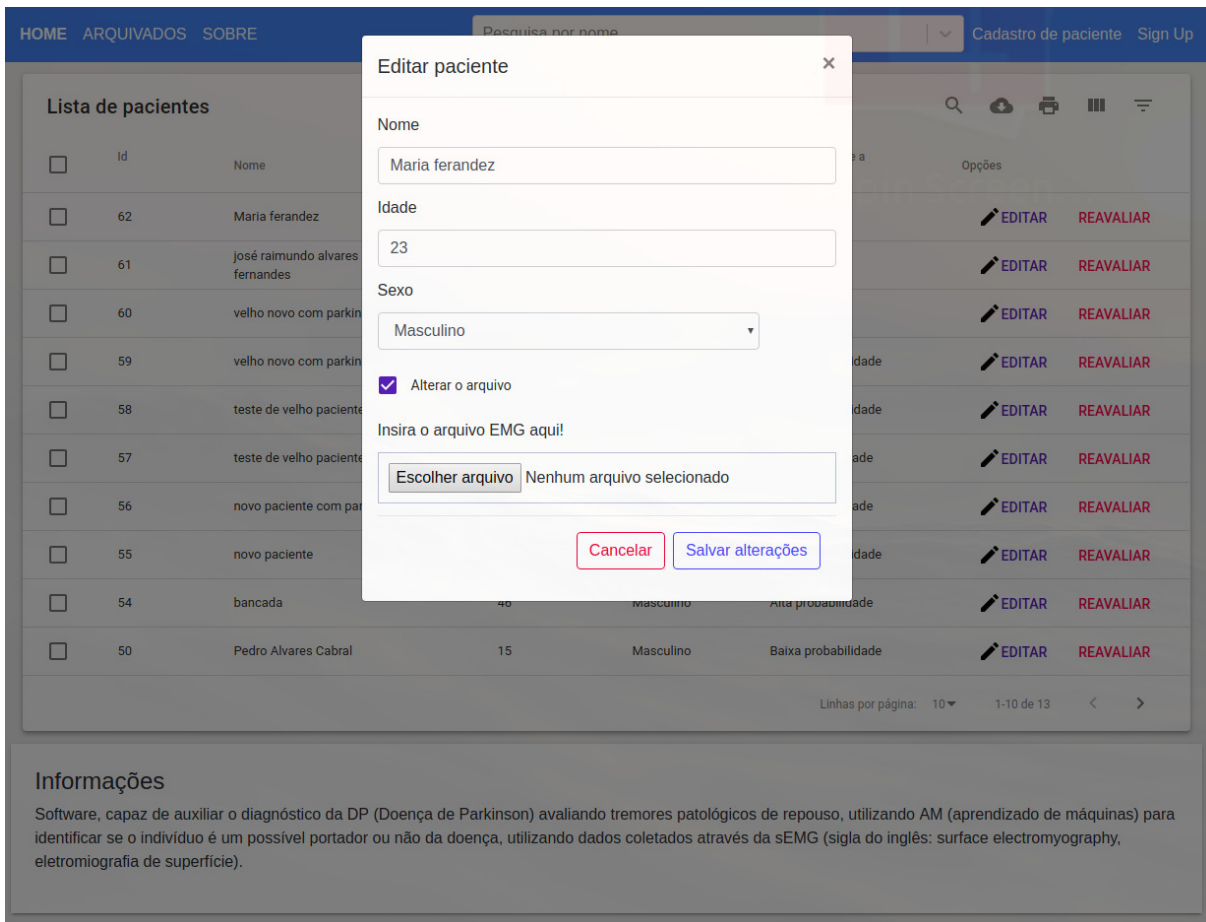


Figura 18 – Editar dados dos pacientes (Imagem dos autores).

8.2.3.6 Pesquisar pacientes

Também é possível executar outras opções como *download* da tabela, impressão e exibir colunas. Como pode ser observado na Figura 20.

8.2.3.7 Pagina arquivados

Para exibir os itens arquivados, clica-se na aba **ARQUIVADOS**, como na Figura 22.

8.2.3.8 Página sobre

A página sobre exibe as informações referentes ao trabalho, Figura 23.

8.3 Passos futuros e considerações finais

Conseguiu-se atingir os objetivos propostos, provando que é possível classificar com uma boa acurácia a DP utilizando AM. Com aproximadamente 79% de precisão, isto considerando o baixo nível de amostras, 36 no total. Além disto, foi desenvolvida

Lista de pacientes

<input type="checkbox"/>	Id	Nome	Idade	Sexo	Resultado sobre a DP	Opções
<input type="checkbox"/>	62	Maria ferandez	23	Masculino	AVALIAR	EDITAR REAVALIAR
<input type="checkbox"/>	61	josé raimundo alvares fernandes	23	Masculino	AVALIAR	EDITAR REAVALIAR
<input type="checkbox"/>	60	velho novo com parkinsom	23	Masculino	AVALIAR	EDITAR REAVALIAR
<input type="checkbox"/>	59	velho novo com parkinsom	23	Masculino	Baixa probabilidade	EDITAR REAVALIAR
<input type="checkbox"/>	58	teste de velho paciente	33	Masculino	Baixa probabilidade	EDITAR REAVALIAR
<input type="checkbox"/>	57	teste de velho paciente	33	Masculino	Alta probabilidade	EDITAR REAVALIAR
<input type="checkbox"/>	56	novo paciente com parkinsom	23	Masculino	Alta probabilidade	EDITAR REAVALIAR
<input type="checkbox"/>	55	novo paciente	22	Masculino	Baixa probabilidade	EDITAR REAVALIAR
<input type="checkbox"/>	54	bancada	46	Masculino	Alta probabilidade	EDITAR REAVALIAR
<input type="checkbox"/>	50	Pedro Alvares Cabral	15	Masculino	Baixa probabilidade	EDITAR REAVALIAR

Linhas por página: 10 1-10 de 13

Informações

Software, capaz de auxiliar o diagnóstico da DP (Doença de Parkinson) avaliando tremores patológicos de repouso, utilizando AM (aprendizado de máquinas) para identificar se o indivíduo é um possível portador ou não da doença, utilizando dados coletados através da sEMG (sigla do inglês: surface electromyography, eletromiografia de superfície).

Figura 19 – Avaliar pacientes (Imagem dos autores).

HOME ARQUIVADOS SOBRE Pesquisa por nome Cadastro de paciente Sign Up

jo

<input type="checkbox"/>	Id	Nome	Idade	Sexo	Resultado sobre a DP	Opções
<input type="checkbox"/>	61	josé raimundo alvares fernandes	23	Masculino	AVALIAR	EDITAR REAVALIAR
<input type="checkbox"/>	41	José da silva freitas	33	Masculino	Alta probabilidade	EDITAR REAVALIAR

Linhas por página: 10 1-2 de 2

Informações

Software, capaz de auxiliar o diagnóstico da DP (Doença de Parkinson) avaliando tremores patológicos de repouso, utilizando AM (aprendizado de máquinas) para identificar se o indivíduo é um possível portador ou não da doença, utilizando dados coletados através da sEMG (sigla do inglês: surface electromyography, eletromiografia de superfície).

Figura 20 – Pesquisar itens pelas colunas (Imagem dos autores).

um *software* totalmente operacional, onde um profissional da saúde consegue classificar facilmente uma pessoa em alta ou baixa probabilidade de possuir a DP.

É desejável a realização de mais coletas, com a finalidade de aumentar as amostras

HOME ARQUIVADOS SOBRE
Pesquisa por nome
Cadastro de paciente Sign Up

Lista de pacientes

<input type="checkbox"/>	Id	Nome	Idade	Sexo	Resultado sobre a DP	Opções
<input type="checkbox"/>	63	Ioerana	22	Feminino	AVALIAR	EDITAR REAVALIAR
<input type="checkbox"/>	62	Maria ferandez	23	Feminino	AVALIAR	EDITAR REAVALIAR
<input type="checkbox"/>	61	josé raimundo alvares fernandes	23	Masculino	AVALIAR	EDITAR REAVALIAR
<input type="checkbox"/>	60	velho novo com parkinsom	23	Masculino	AVALIAR	EDITAR REAVALIAR
<input type="checkbox"/>	59	velho novo com parkinsom	23	Masculino	Baixa probabilidade	EDITAR REAVALIAR
<input type="checkbox"/>	58	Luciana Fonseca	33	Feminino	Baixa probabilidade	EDITAR REAVALIAR
<input type="checkbox"/>	57	teste de velho paciente	33	Masculino	Alta probabilidade	EDITAR REAVALIAR
<input type="checkbox"/>	56	novo paciente com parkinsom	23	Masculino	Alta probabilidade	EDITAR REAVALIAR
<input type="checkbox"/>	55	Ana Costa	22	Feminino	Baixa probabilidade	EDITAR REAVALIAR
<input type="checkbox"/>	54	bancada	46	Masculino	Alta probabilidade	EDITAR REAVALIAR

Linhas por página: 10 1-10 de 14

Informações






Software, capaz de auxiliar o diagnóstico da DP (Doença de Parkinson) avaliando tremores patológicos de repouso, utilizando AM (aprendizado de máquinas) para identificar se o indivíduo é um possível portador ou não da doença, utilizando dados coletados através da SEMG (sigla do inglês: surface electromyography, eletromiografia de superfície).

Figura 21 – *Download* da tabela, impressão e exibir colunas (Imagem dos autores).

de treinamento e consequentemente aumentar a acurácia na classificação da DP, também é desejável a realização de uma anamnese mais detalhada. Isso em passos futuros.

[HOME](#) **ARQUIVADOS** [SOBRE](#) [Cadastro de paciente](#) [Sign Up](#)

Pacientes arquivados

Id	Nome	Idade	Sexo	Resultado sobre a DP	Opções
53	novo	45	Masculino	Alta probabilidade	DESARQUIVAR
52	novo	45	Masculino	Baixa probabilidade	DESARQUIVAR
51	José de anchieta	45	Masculino	Baixa probabilidade	DESARQUIVAR
49	josé raimundo alvares fernandes	3	Masculino	Alta probabilidade	DESARQUIVAR
48	brilhamuito	12	Masculino	Alta probabilidade	DESARQUIVAR
47	ronaldo	22	Masculino	Alta probabilidade	DESARQUIVAR
46	novo	14	Masculino	Baixa probabilidade	DESARQUIVAR
45	testeUltimo	12	Masculino	Alta probabilidade	DESARQUIVAR
44	ultimo	33	Masculino	Baixa probabilidade	DESARQUIVAR
40	maisumtest	33	Masculino	Baixa probabilidade	DESARQUIVAR

Linhas por página: 10 1-10 de 49 < >

Informações

Software, capaz de auxiliar o diagnóstico da DP (Doença de Parkinson) avaliando tremores patológicos de repouso, utilizando AM (aprendizado de máquinas) para identificar se o indivíduo é um possível portador ou não da doença, utilizando dados coletados através da sEMG (sigla do inglês: surface electromyography, eletromiografia de superfície).

Figura 22 – *Download* da tabela, impressão e exibir colunas (Imagem dos autores).

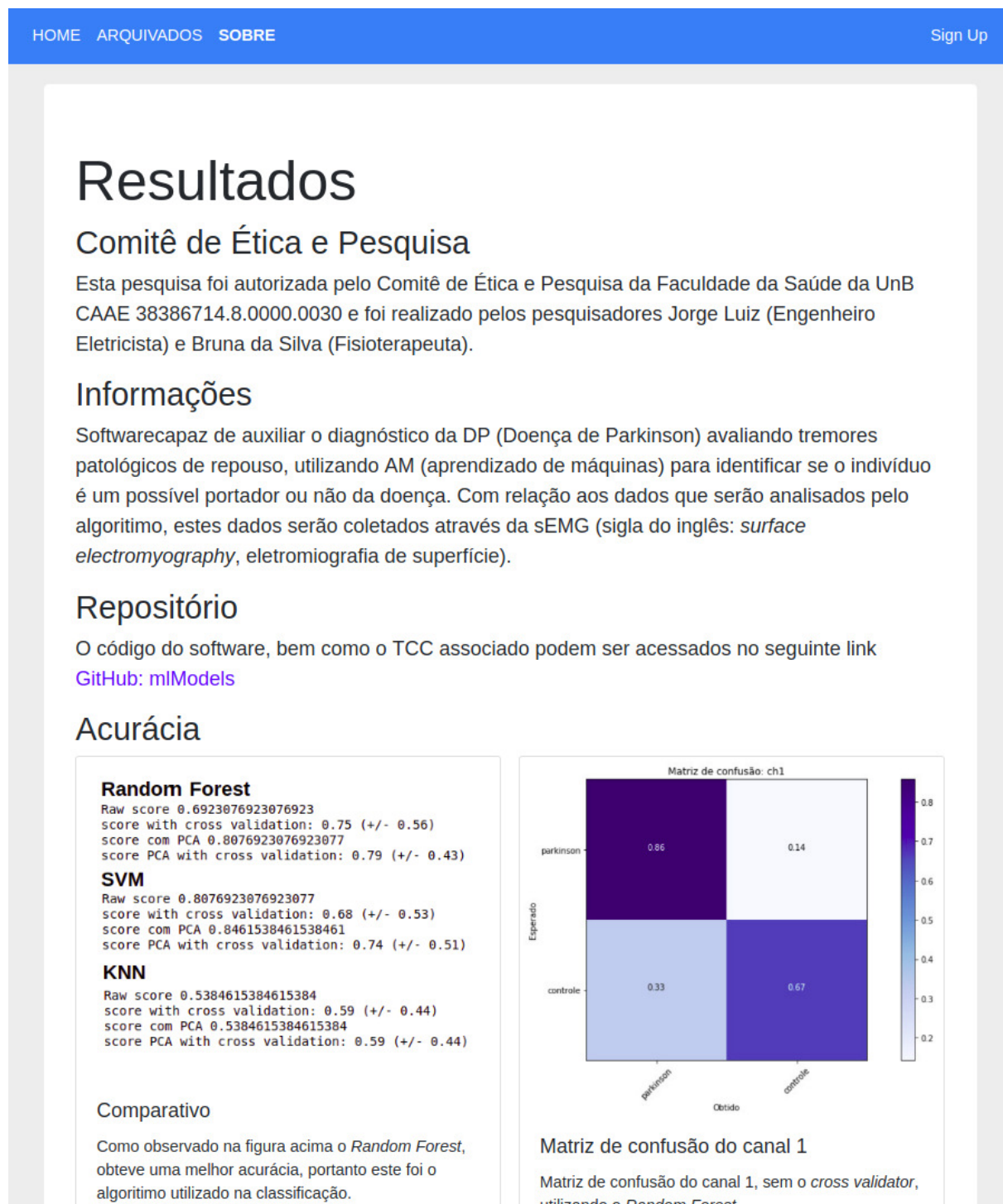


Figura 23 – informações sobre o projeto, na página Sobre (Imagem dos autores).

Referências

- ABADI, M. et al. Tensorflow: A system for large-scale machine learning. In: *OSDI*. [S.l.: s.n.], 2016. v. 16, p. 265–283. Citado na página 32.
- AI, L.; WANG, J.; YAO, R. Classification of parkinsonian and essential tremor using empirical mode decomposition and support vector machine. *Digital Signal Processing*, Elsevier, v. 21, n. 4, p. 543–550, 2011. Citado na página 52.
- AMAZON. *Amazon Machine Learning Developer Guide*. [S.l.]: Amazon Web Services, Inc. and/or its affiliates, 2018. Citado 4 vezes nas páginas 27, 55, 56 e 57.
- AMIT, Y.; GEMAN, D. Shape quantization and recognition with randomized trees. *Neural computation*, MIT Press, v. 9, n. 7, p. 1545–1588, 1997. Citado na página 39.
- BARBOSA, E.; SALLEM, F. Parkinsons disease–diagnosis. *REVISTA NEUROCIÊNCIAS*, v. 13, n. 3, p. 158–165, 2005. Citado na página 44.
- BASTOS, D. G.; NASCIMENTO, P. S.; LAURETTO, M. S. Proposta e análise de desempenho de dois métodos de seleção de características para random forests. *IX Simpósio Brasileiro de Sistemas de Informação*, p. 49–60, 2013. Citado na página 39.
- BATISTA, G. E. d. A. P. et al. *Pré-processamento de dados em aprendizado de máquina supervisionado*. Tese (Doutorado) — Universidade de São Paulo, 2003. Citado na página 49.
- BLUM, A. L.; LANGLEY, P. Selection of relevant features and examples in machine learning. *Artificial intelligence*, Elsevier, v. 97, n. 1-2, p. 245–271, 1997. Citado na página 57.
- BORGES, H. B.; NIEVOLA, J. *Redução de Dimensionalidade em Bases de Dados de Expressão Gênica*. Tese (Doutorado) — Dissertação de Mestrado, PPGIa-PUCPR, 2006. Citado na página 30.
- BOTELHO, A. L. et al. Avaliação eletromiográfica de assimetria dos músculos mastigatórios em sujeitos com oclusão normal. *Revista da Faculdade de Odontologia-UPF*, v. 13, n. 3, 2010. Citado na página 45.
- BRAGA, A. d. P.; CARVALHO, A.; LUDERMIR, T. B. *Redes neurais artificiais: teoria e aplicações*. [S.l.]: Livros Técnicos e Científicos, 2000. Citado na página 35.
- BREIMAN, L. Bagging predictors. *Machine learning*, Springer, v. 24, n. 2, p. 123–140, 1996. Citado 2 vezes nas páginas 39 e 40.
- BREIMAN, L. Random forests. *Machine learning*, Springer, v. 45, n. 1, p. 5–32, 2001. Citado 2 vezes nas páginas 39 e 40.
- BREIMAN, L. *Classification and regression trees*. [S.l.]: Routledge, 2017. Citado na página 40.

- BRIGHAM, E. O.; BRIGHAM, E. O. *The fast Fourier transform and its applications*. [S.l.]: prentice Hall Englewood Cliffs, NJ, 1988. v. 448. Citado na página 30.
- CAELEN, O. A bayesian interpretation of the confusion matrix. *Annals of Mathematics and Artificial Intelligence*, Springer, v. 81, n. 3-4, p. 429–450, 2017. Citado na página 33.
- CAMARA, C. et al. Resting tremor classification and detection in parkinson's disease patients. *Biomedical Signal Processing and Control*, Elsevier, v. 16, p. 88–97, 2015. Citado 2 vezes nas páginas 25 e 51.
- CAMP, R. C. Benchmarking: o caminho da qualidade total. *São Paulo: Pioneira*, v. 3, 1998. Citado na página 32.
- CHANDRASHEKAR, G.; SAHIN, F. A survey on feature selection methods. *Computers & Electrical Engineering*, Elsevier, v. 40, n. 1, p. 16–28, 2014. Citado na página 30.
- CHRISTIE, T. Django rest framework. *Home-Django REST Framework*, 2011. Citado na página 54.
- COSTA, F. P. da; DIAFÉRIA, G.; BEHLAU, M. Aspectos comunicativos e enfrentamento da disfonia em pacientes com doença de parkinson. 2016. Citado 2 vezes nas páginas 25 e 43.
- EFTAXIAS, K. et al. Detection of parkinson's tremor from emg signals; a singular spectrum analysis approach. In: IEEE. *Digital Signal Processing (DSP), 2015 IEEE International Conference on*. [S.l.], 2015. p. 398–402. Citado 3 vezes nas páginas 25, 43 e 47.
- EVGENIOU, T.; PONTIL, M. Support vector machines: Theory and applications. In: SPRINGER. *Advanced Course on Artificial Intelligence*. [S.l.], 1999. p. 249–257. Citado 2 vezes nas páginas 37 e 38.
- Facebook Inc. *React: A JavaScript library for building user interfaces*. 2018. <<https://reactjs.org/>>. Accessed: 2018-05-14. Citado na página 54.
- FAWAGREH, K.; GABER, M. M.; ELYAN, E. Random forests: from early developments to recent advancements. *Systems Science & Control Engineering: An Open Access Journal*, Taylor & Francis, v. 2, n. 1, p. 602–609, 2014. Citado na página 39.
- GAGO, M. et al. Manual para pessoas com parkinson. *APDPk-Assoc. Port. Doentes Park.*, 2014. Citado 2 vezes nas páginas 25 e 44.
- GÉRON, A. *Hands-on machine learning with Scikit-Learn and TensorFlow: concepts, tools, and techniques to build intelligent systems*. [S.l.]: "O'Reilly Media, Inc.", 2017. Citado 7 vezes nas páginas 28, 29, 30, 31, 55, 56 e 58.
- GÜLER, N. F.; KOÇER, S. Classification of emg signals using pca and fft. *Journal of Medical Systems*, Springer, v. 29, n. 3, p. 241–250, 2005. Citado na página 51.
- HAYKIN, S. Neural networks: A comprehensive foundation, vol. 2. *Segundo, Pentrice Hall. España*, 1999. Citado na página 35.

- HO, T. K. Random decision forests. In: IEEE. *Document analysis and recognition, 1995., proceedings of the third international conference on*. [S.l.], 1995. v. 1, p. 278–282. Citado 2 vezes nas páginas 28 e 39.
- HOSCH, W. L. Corrected parenthetical definition of genetic algorithm to distinguish between symbols, which are bred, and algorithms, which are used to cull the symbols. 2009. Citado na página 27.
- HUANG, C.; DAVIS, L.; TOWNSHEND, J. An assessment of support vector machines for land cover classification. *International Journal of remote sensing*, Taylor & Francis, v. 23, n. 4, p. 725–749, 2002. Citado na página 36.
- JAMES, G. et al. *An introduction to statistical learning*. [S.l.]: Springer, 2013. v. 112. Citado na página 33.
- JANKOVIC, J. Parkinson's disease: clinical features and diagnosis. *Journal of Neurology, Neurosurgery & Psychiatry*, BMJ Publishing Group Ltd, v. 79, n. 4, p. 368–376, 2008. Citado na página 46.
- JOLLIFFE, I. Principal component analysis. In: *International encyclopedia of statistical science*. [S.l.]: Springer, 2011. p. 1094–1096. Citado na página 30.
- KATSIS, C. D. et al. A novel method for automated emg decomposition and muap classification. *Artificial Intelligence in Medicine*, Elsevier, v. 37, n. 1, p. 55–64, 2006. Citado na página 49.
- KIRK, M. *Thoughtful machine learning: A test-driven approach*. [S.l.]: "O'Reilly Media, Inc.", 2014. Citado 2 vezes nas páginas 28 e 29.
- KOHAVI, R. Glossary of terms. *Machine Learning*, v. 30, p. 271–274, 1998. Citado 2 vezes nas páginas 25 e 27.
- KUGLER, P. et al. Automated classification of parkinson's disease and essential tremor by combining electromyography and accelerometer signals. *Basal Ganglia*, Elsevier, v. 3, n. 1, p. 61, 2013. Citado 2 vezes nas páginas 25 e 52.
- LIU, J. Feature dimensionality reduction for myoelectric pattern recognition: A comparison study of feature selection and feature projection methods. *Medical Engineering and Physics*, Elsevier, v. 36, n. 12, p. 1716–1720, 2014. Citado na página 52.
- LOCONSOLE, C. et al. A model-free technique based on computer vision and semg for classification in parkinson's disease by using computer-assisted handwriting analysis. *Pattern Recognition Letters*, Elsevier, 2018. Citado 2 vezes nas páginas 25 e 52.
- LORENA, A. C.; CARVALHO, A. C. de. Uma introdução às support vector machines. *Revista de Informática Teórica e Aplicada*, v. 14, n. 2, p. 43–67, 2007. Citado 2 vezes nas páginas 28 e 35.
- MAIA, J. *Doença de Parkinson*. 2018. <<http://fisioterapiajoaomaia.blogspot.com.br/2012/10/doenca-de-parkinson-doenca-de-parkinson.html>>. Accessed: 2018-4-30. Citado 2 vezes nas páginas 15 e 43.

- MCKINNEY, W. *Python for data analysis: Data wrangling with Pandas, NumPy, and IPython*. [S.l.]: "O'Reilly Media, Inc.", 2012. Citado na página 32.
- MITCHELL, T. M. et al. Machine learning. 1997. *Burr Ridge, IL: McGraw Hill*, v. 45, n. 37, p. 870–877, 1997. Citado na página 27.
- MURTHY, S. K. Automatic construction of decision trees from data: A multi-disciplinary survey. *Data mining and knowledge discovery*, Springer, v. 2, n. 4, p. 345–389, 1998. Citado na página 28.
- NASCIMENTO, G. K. B. O. et al. Surface electromyography of the masseter muscle during chewing: a systematic review. *Revista CEFAC*, SciELO Brasil, v. 14, n. 4, p. 725–731, 2012. Citado na página 47.
- NOBLE, W. S. et al. Support vector machine applications in computational biology. *Kernel methods in computational biology*, Cambridge, p. 71–92, 2004. Citado na página 35.
- ORTOLAN, R. *ESTUDO E AVALIAÇÃO DE TÉCNICAS DE PROCESSAMENTO DO SINAL MIOELÉTRICO PARA O CONTROLE DE SISTEMAS DE REABILITAÇÃO*. 2002. Citado 2 vezes nas páginas 15 e 47.
- PAMPLONA, L. d. A.; MATTOS, J. P. d. et al. Demência na doença de parkinson: avaliação crítica da literatura. *Arq Neuropsiquiatr*, v. 54, n. 4, p. 687–90, 1996. Citado na página 44.
- PEDREGOSA, F. et al. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, v. 12, p. 2825–2830, 2011. Citado 9 vezes nas páginas 15, 32, 33, 34, 35, 36, 37, 39 e 40.
- PEIXINHO, A.; AZEVEDO, A. L.; SIMÕES, R. M. Alterações neuropsiquiátricas da doença de parkinson. *Psilogos*, v. 3, n. 2, p. 12–30, 2006. Citado na página 44.
- PERNAMBUCO, L. de A. et al. A eletromiografia de superfície nos periódicos nacionais em fonoaudiologia. *Revista CEFAC*, SciELO Brasil, v. 12, n. 4, 2010. Citado 2 vezes nas páginas 25 e 45.
- PETERNELLA, F. M. N.; MARCON, S. S. Descobrimos a doença de parkinson: impacto para o parkinsoniano e seu familiar. *Revista Brasileira de Enfermagem*, Associação Brasileira de Enfermagem, v. 62, n. 1, 2009. Citado na página 44.
- PETERSON, L. E. K-nearest neighbor. *Scholarpedia*, v. 4, n. 2, p. 1883, 2009. Citado 2 vezes nas páginas 41 e 42.
- PHINYOMARK, A.; PHUKPATTARANONT, P.; LIMSAKUL, C. Feature reduction and selection for emg signal classification. *Expert Systems with Applications*, Elsevier, v. 39, n. 8, p. 7420–7431, 2012. Citado na página 50.
- PONTIL, M.; VERRI, A. Support vector machines for 3d object recognition. *IEEE transactions on pattern analysis and machine intelligence*, IEEE, v. 20, n. 6, p. 637–646, 1998. Citado na página 35.
- RASCHKA, S. *Python machine learning*. [S.l.]: Packt Publishing Ltd, 2015. Citado 4 vezes nas páginas 15, 33, 34 e 35.

- REAZ, M. B. I.; HUSSAIN, M.; MOHD-YASIN, F. Techniques of emg signal analysis: detection, processing, classification and applications. *Biological procedures online*, v. 8, n. 1, p. 11, 2006. Citado na página 45.
- ROSA, I. da G. *DETECÇÃO AUTOMÁTICA DE CONTRAÇÃO MUSCULAR UTILIZANDO A ELETROMIOGRAFIA DE SUPERFÍCIE*. Tese (Doutorado) — UNIVERSIDADE FEDERAL DO RIO DE JANEIRO, 2005. Citado 2 vezes nas páginas 45 e 46.
- SABINO, M. *Unidades motoras e a Musculação*. 2017. <https://www.sas.com/pt_br/insights/analytics/machine-learning.html>. Accessed: 2018-06-10. Citado 2 vezes nas páginas 15 e 47.
- SAITO, T. C. A doença de parkinson e seus tratamentos: uma revisão bibliográfica. *Paper Specialization Londrina: Centro Universitário Filadélfia*, 2011. Citado na página 44.
- SAMUEL, A. L. Some studies in machine learning using the game of checkers. 1959. Citado na página 27.
- SAS. *Machine Learning. O que é e qual sua importância?* 2018. <<http://tafitness.net/unidades-motoras-musculacao/>>. Accessed: 2018-04-10. Citado na página 27.
- SCHMITZ, J. et al. Unchanged muscle fiber conduction velocity relates to mild acidosis during exhaustive bicycling. *European journal of applied physiology*, Springer, v. 112, n. 5, p. 1593–1602, 2012. Citado na página 47.
- SENDERS, J. T. et al. Machine learning and neurosurgical outcome prediction: A systematic review. *World Neurosurgery*, Elsevier, 2017. Citado na página 32.
- SOTO, M. F.; SEGURA-TORRES, J. Estructura y función de los receptores acetilcolina de tipo muscarínico y nicotínico. *Rev Mex Neuroci*, v. 6, n. 4, p. 315–326, 2005. Citado na página 45.
- SUTHERLAND, J. *Scrum: a arte de fazer o dobro do trabalho na metade do tempo*. [S.l.]: Leya, 2016. Citado na página 53.
- TECHOPEDIA. *Minimum Viable Product (MVP) definition*. 2018. <<https://www.techopedia.com/definition/27809/minimum-viable-product-mvp>>. Accessed: 2018-11-12. Citado na página 54.
- TREVOR, H.; ROBERT, T.; JH, F. *The elements of statistical learning: data mining, inference, and prediction*. [S.l.]: New York, NY: Springer, 2009. Citado na página 28.
- VAPNIK, V. *Statistical learning theory*. 1998. [S.l.]: Wiley, New York, 1998. v. 3. Citado 2 vezes nas páginas 36 e 38.
- VORONOI, G. G. voronoi, j. reine angew. math. 133, 97 (1907). *J. Reine Angew. Math.*, v. 133, p. 97, 1907. Citado 3 vezes nas páginas 15, 41 e 42.
- WAHBA, G. *Spline models for observational data*. [S.l.]: Siam, 1990. v. 59. Citado 2 vezes nas páginas 36 e 37.
- WALD, A. *Statistical decision functions*. Wiley, 1950. Citado na página 37.

YOUSEFI, J.; HAMILTON-WRIGHT, A. Characterizing emg data using machine-learning tools. *Computers in biology and medicine*, Elsevier, v. 51, p. 1–13, 2014. Citado 5 vezes nas páginas [15](#), [46](#), [49](#), [50](#) e [51](#).

ZARZUR, A. P. et al. Laryngeal electromyography and acoustic voice analysis in parkinson's disease: a comparative study. *Brazilian journal of Otorhinolaryngology*, Elsevier, v. 76, n. 1, p. 40–43, 2010. Citado na página [44](#).

Apêndices

APÊNDICE A – Documento de Arquitetura

A.1 Introdução

Este documento descreve a arquitetura e tecnologias da solução de *software*, utilizadas no desenvolvimento do sistema de auxílio na detecção da doença de Parkinson avaliando o Tremor em Repouso. Desenvolvido pelos alunos de Trabalho de Conclusão de Curso 2 da Universidade de Brasília.

Essa solução de *software* trata-se de uma aplicação *web*, na qual pode-se facilmente ser migrada para uma aplicação *mobile* em projetos futuros.

A.2 Tecnologias

A.2.1 SPA

SPA (sigla do inglês: *Single Page Application*, Aplicativo de página única), consiste em uma aplicação *web* em uma única página, fornecendo uma experiência de usuário semelhante a um aplicativo Desktop, sem recarregamentos da página.

O SPA foi adotado devido a sua vantagem de performance no consumo de dados e na experiência do usuário.

A.2.2 REST

Baseado no protocolo de comunicação de rede HTTP, sendo simples, rápida e com uma fácil comunicação entre clientes e servidores. Funciona com objetos em formato JSON, e os métodos definidos no protocolo HTTP (POST, GET, PUT e DELETE) possibilitando assim uma eficiente e rápida troca de informação.

Foi adotado devido a seu leve consumo de dados, além de possuir uma fácil implementação.

A.3 Representação da Arquitetura

A arquitetura se baseia em dois modelos arquiteturais, REST e SPA, interligados com um modelo de aprendizado de máquinas. No caso a *Random Forest* ou o SVM, como observado na Figura 24. Deste modo a aplicação possui três pacotes:

- *Machine Learning models*;

- API;
- Web.

A.4 Visão Lógica

A.4.1 Diagrama de Pacotes

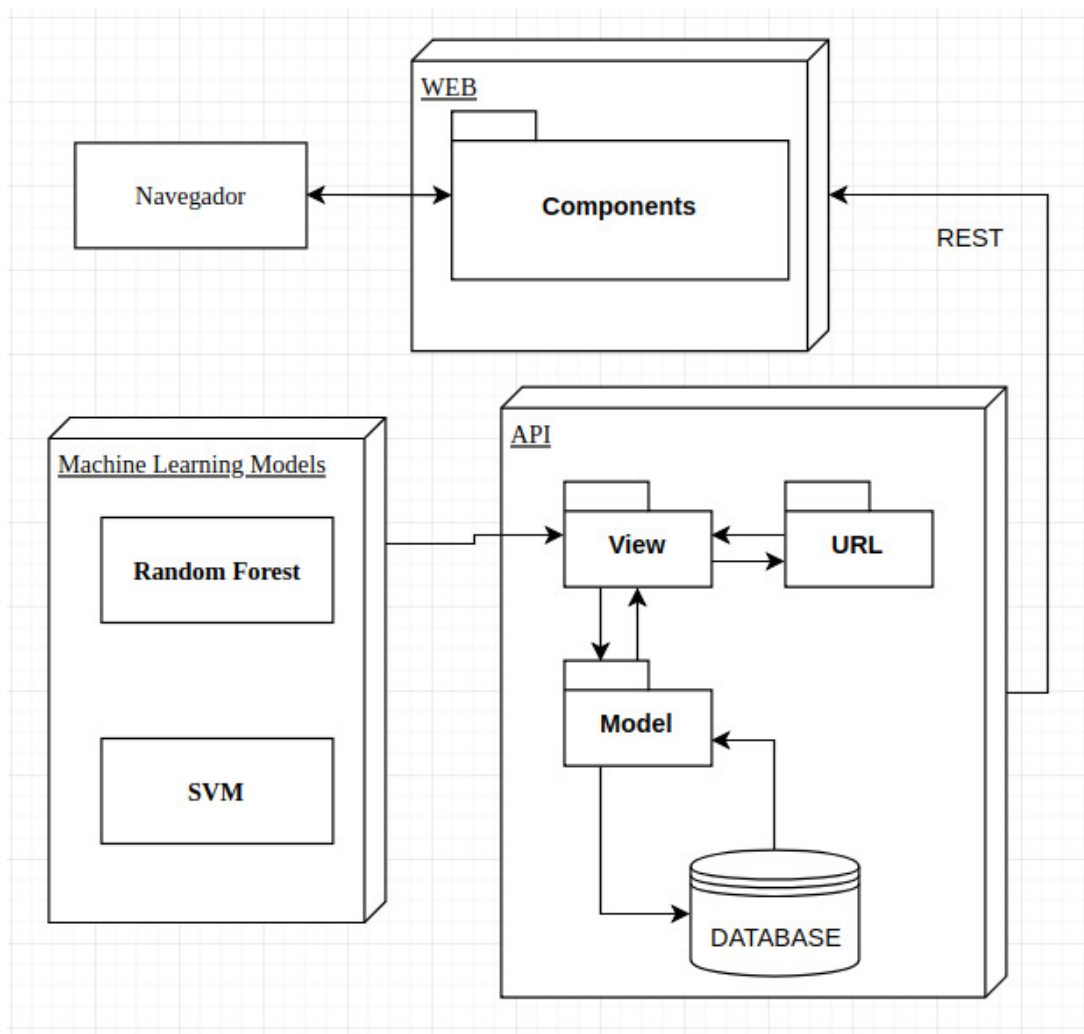


Figura 24 – Diagrama de pacotes.

A.4.1.1 Machine Learning models

Consiste no modelo já treinado e validado do algoritmo *Random Forest*, sendo que este é armazenado no banco de dados *PickleDB*. Deste modo, a comunicação é somente de leitura, com o modelo sendo enviado por completo para a API.

A.4.1.2 API

A API executa todas as manipulações envolvendo os dados, como o pré-processamento do arquivo enviado pelo usuário, e classificando estes com o modelo salvo. Além de salvar os dados no banco de dados e enviá-los quando requisitados, a API foi desenvolvida utilizando os padrões de projeto definidos pelo Django-Rest-Framework, possuindo o maior esforço na *View*.

No caso as seguintes classes foram desenvolvidas, sendo elas:

1. *InsertFeature*: realiza o cadastro de um paciente.
2. *ListAllPatients*: lista todos os pacientes, retornando-os em um objeto JSON.
3. *DetailPatient*: retorna os detalhes relativos a um único paciente.
4. *UpdatePatient*: possibilita a edição de um paciente já cadastrado.
5. *CLS*: possui apenas um método GET, que recebe um *id* de um paciente. Com auxílio de outros métodos carrega o arquivo sEMG associado a este paciente, realiza o pré-processamento aplicando a FFT, carrega o modelo e por fim retorna o resultado da classificação.

A.4.1.3 Web

Responsável pela exibição dos dados, desenvolvido em *React*. Possui somente três componentes, cadastrado, editar e home.

A.4.2 Diagrama de Classes

A seguir na Figura 25, encontra-se o diagrama de classes do projeto.

A.5 Restrições e Metas Arquiteturais

A.5.1 Restrições

- Client-Server: O sistema deverá rodar em um servidor, podendo ser instalado em uma rede local.

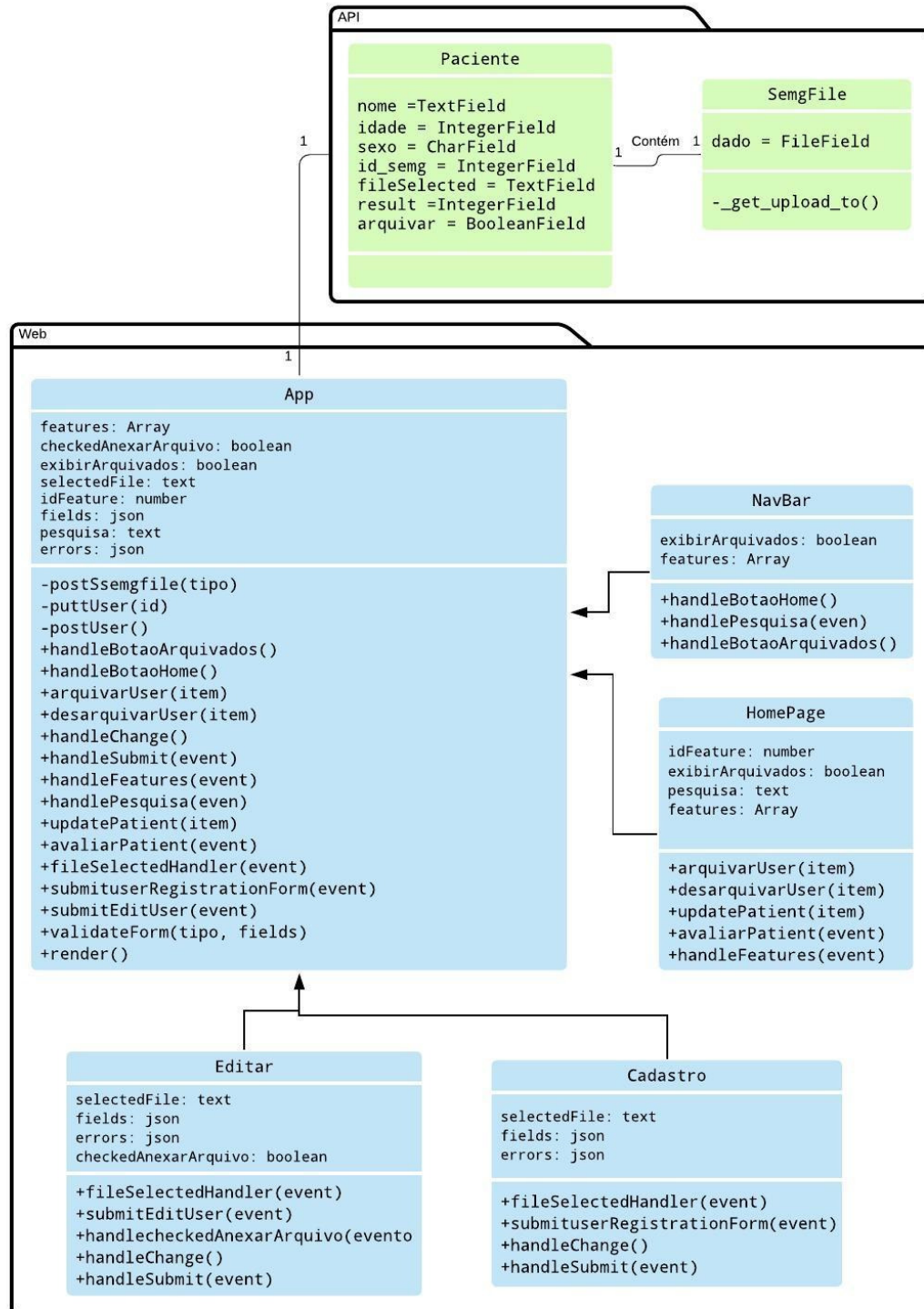


Figura 25 – Diagrama de Classes.

APÊNDICE B – Documento de Visão

B.1 Introdução

Este documento tem como propósito descrever todos os requisitos de um sistema capaz de realizar um diagnóstico preliminar da doença de Parkinson em pacientes com suspeita da mesma, denominado Detector preliminar da Doença de Parkinson (DPDP). Serão explicitados os aspectos inerentes ao sistema, visando uma melhor definição do projeto de forma simplificada. O documento está organizado em 8 tópicos que visam uma melhor compreensão do leitor, sendo eles: introdução, posicionamento, descrição da parte interessada e do usuário, visão geral do produto, recursos do produto, restrições. Em todos os tópicos são abordados pontos cruciais para o desenvolvimento do projeto.

B.1.1 Problema

Atualmente o diagnóstico da doença de Parkinson é feito avaliando-se a história do paciente, o seu exame neurológico, resposta à terapia dopaminérgica e avaliação dos sinais obtidos por dispositivos sEMG, esses métodos, apesar de eficazes, ainda demandam um certo tempo para alcançar o diagnóstico concreto da doença. Portanto o desenvolvimento de um sistema utilizando aprendizagem de máquina torna-se um viés mais eficaz para auxiliar nessas etapas do diagnóstico da doença de Parkinson.

B.1.2 Escopo

O DPDP permite que o médico (usuário) possa fazer uma avaliação preliminar prévia do estado do paciente quanto a presença ou não da doença de Parkinson inserindo dados obtidos por um dispositivo sEMG. O sistema também permite que o médico consulte no banco de dados os pacientes que já seu diagnóstico preliminar.

B.2 Posicionando

B.2.1 Oportunidade de Negócios

Apesar dos meios para o diagnóstico da doença de Parkinson serem considerados suficientes, o desenvolvimento de um sistema capaz de melhorar ainda mais a eficiência desse diagnóstico, utilizando os conceitos do aprendizado de máquina, ainda tendo em vista que não existem outros meios semelhantes a esse sistema para aferir um diagnóstico

preliminar da doença, pode significar um passo a mais nas pesquisas sobre novos diagnósticos, não apenas da doença de Parkinson, mas também de outros distúrbios neurológicos.

B.2.2 Instrução do Problema

A Tabela 3 descreve brevemente a oportunidade de negócios que é tratada por este projeto segundo o problema no qual o mesmo se dedica a solucionar em uma instrução geral resumida.

Tabela 3 – Instrução do Problema

O problema	Baixa eficiência do diagnóstico da doença de Parkinson
Afeta	Os pacientes
Cujo o impacto é	Um possível diagnóstico tardio da doença
Uma boa solução seria	Sistema automatizado capaz fornecer um diagnóstico prévio da doença de Parkinson

B.3 Descrições da Parte Interessada e do Usuário

Para fornecer produtos e serviços que atendam às necessidades das partes interessadas e dos usuários é necessário identificar e envolver todas essas partes dentro do processo de definição dos requisitos. Essa seção se dedica a apresentar e detalhar os perfis dos interessados no projeto e usuários do sistema.

B.3.1 Resumo da Parte Interessada

A Tabela 4 possui um breve resumo dos interessados no projeto, descrevendo-os e apresentando suas respectivas responsabilidades dentro do mesmo.

Tabela 4 – Resumo da Parte Interessada

Nome	Descrição	Responsabilidade
Orientandos	Estudantes matriculados na disciplina de TCC 2	Desenvolvimento da solução de <i>software</i> e pesquisa.
Orientadores	Professor(a) orientador(a) e pesquisadores	Orientadores e avaliadores que darão suporte a respeito das pesquisas e escrita do TCC assim como no desenvolvimento do sistema.

B.3.2 Resumo do Usuário

A Tabela 5 apresenta um breve resumo do usuário do sistema com sua descrição e seu papel desempenhado na operação do mesmo.

Tabela 5 – Resumo do Usuário

Nome	Descrição	Responsabilidade
Médicos	Pessoas capacitadas da área da saúde	Inserir os dados dos pacientes, verificar o diagnóstico e consultar os dados dos pacientes no banco

B.3.3 Perfis das Partes Interessadas

Essa seção descreve todas as partes interessadas no sistema de maneira mais detalhada demonstrando seu grau de envolvimento no projeto, critérios para sucesso do mesmo e suas responsabilidades dentro do projeto.

B.3.3.1 Orientandos

A Tabela 6 detalha o perfil dos orientandos do projeto.

Tabela 6 – Perfil dos Orientandos

Representantes	Flávio da Costa Paixão, Jônatas Lennon Lima Costa
Descrição	Desenvolvedores do sistema e gerenciamento do projeto
Tipo	Estudantes matriculados na disciplina de TCC 2
Responsabilidades	Desenvolvimento da solução de <i>software</i> e pesquisa
Crítérios de Sucesso	Manter os prazos estabelecidos sem atraso, e gerenciar a qualidade do <i>software</i> em desenvolvimento
Envolvimento	Alto

B.3.3.2 Orientadores

A Tabela 7 detalha o perfil dos orientadores do projeto.

Tabela 7 – Perfil dos Orientadores

Representantes	Dra. Lourdes Mattos Brasil, Bel. Ithallo Junior Alves Guimarães
Descrição	Product Owner
Tipo	Orientadores e avaliadores que darão suporte a respeito das pesquisas e escrita do TCC assim como no desenvolvimento do sistema.
Responsabilidades	Orientar e avaliar
Crítérios de Sucesso	Olhar crítico na avaliação do projeto
Envolvimento	Alto

B.3.4 Perfis do Usuário

Essa seção descreve o usuário do sistema e assim como os perfis das partes interessadas apresentadas na Tabela 6 e Tabela 7 também descreve de maneira mais detalhada demonstrando seu grau de envolvimento no projeto, critérios para sucesso do mesmo e suas responsabilidades dentro do projeto.

B.3.4.1 Médicos

A Tabela 8 descreve o perfil do médico, usuário principal do sistema.

Tabela 8 – Perfil dos Médicos

Representantes	Médicos
Descrição	Pessoa capacitada para verificar o diagnóstico da doença de Parkinson em seus pacientes
Tipo	Profissionais que tem o interesse em utilizar o sistema para validar um diagnóstico preliminar da doença de parkinson em seus pacientes.
Responsabilidades	Verificar o diagnóstico do paciente através do sistema
Critérios de Sucesso	Manter-se ciente de que o diagnóstico é preliminar e serve apenas como ferramenta de auxílio para a identificação de alguns padrões que podem indicar a presença da doença no paciente
Envolvimento	Alto

B.4 Descrição da Solução

B.4.1 Perspectiva do Produto

O sistema tem como objetivo utilizar os sinais obtidos por um dispositivo sEMG para realizar um diagnóstico preliminar da doença de Parkinson utilizando um modelo de aprendizado de máquina.

B.4.2 Resumo dos recursos

A Tabela 9 Apresenta os principais recursos do sistema.

Tabela 9 – Resumo dos recursos

Benefício para o cliente	Recursos de suporte
Automatizar o processo de diagnóstico preliminar	O sistema permite a inclusão dos dados gerados pelos dispositivos sEMG para a verificação da presença da doença
Manter registro de pacientes no banco de dados	O sistema possibilita a consulta de pacientes já diagnosticados preliminarmente

B.4.3 Recursos do Produto

O sistema DPDP oferece as seguintes funcionalidades:

- Incluir os dados do paciente

- Realizar diagnóstico preliminar da doença de Parkinson
- Consultar dados dos pacientes já diagnosticados preliminarmente
- Arquivar dados dos pacientes
- Editar dados dos pacientes

B.4.4 Requisitos Funcionais

A Tabela 10 apresenta todos os requisitos que um o sistema atende.

Tabela 10 – Requisitos Funcionais

ID	Descrição
RF01	Cadastrar paciente
RF02	Editar paciente
RF03	Consultar paciente
RF04	Arquivar paciente
RF05	Baixar dados do paciente
RF06	Imprimir lista de paciente
RF07	Fazer upload de arquivos
RF08	Diagnóstico preliminar

Anexos

ANEXO A – Protocolo - Grupo de Pesquisa Sobre Complicações da Doença de Parkinson



UNIVERSIDADE DE BRASÍLIA - UnB
Faculdade do Gama - FGA

Protocolo - Grupo de Pesquisa Sobre Complicações da Doença de Parkinson

Pesquisadores:
Jorge Luiz - Engenheiro Eletricista
Bruna da Silva - Fisioterapeuta

Orientadora: Prof^ª Dr^ª Lourdes Mattos Brasil
Prof^ª Dr^ª Vera Regina F. da Silva Marães

Introdução

A doença de Parkinson é uma doença neurodegenerativa de etiologia idiopática tendo a idade como o principal fator de risco, porém não o único. Caracteriza-se bioquimicamente pela alfa sinucleína positiva, presença de corpos de Lewis e perda da dopamina na substância negra compacta. Os sintomas da doença são a Bradicinesia, instabilidade postural, tremor ao repouso.

Objetivo do Estudo

A doença de Parkinson embora muito bem elucidada na literatura ainda não possui uma metodologia de alta precisão para o estadiamento da Doença de Parkinson. Assim, esta pesquisa busca através de sinais eletromiográficos e técnicas de aprendizado de máquina caracterizar e classificar os diferentes níveis da Doença de Parkinson, para então agir como uma metodologia automatizada e precisa que irá auxiliar o profissional a realizar o estadiamento da doença. Esta pesquisa foi aprovada pelo Comitê de Ética e Pesquisa da Faculdade da Saúde da Universidade de Brasília CAAE (38386714.8.0000.0030)

Orientações Básicas

O estudo será um estudo piloto onde o objetivo é um n de 20 para grupo controle e de 20 para grupo experimental. O grupo controle será composto por idosos acima de 60 anos, lúcidos e orientados no espaço e tempo, sem problemas circulatórios complexos, sem histórico de AVC com sistema osteomioarticular intacto. O grupo experimental deverá abranger indivíduos portadores da doença de Parkinson, sem outra doença neurodegenerativa, praticantes ou não de atividades físicas. Cada paciente deverá comparecer ao local uma vez, ou seja, um encontro.

Encontro

O paciente será recepcionado com uma cadeira de rodas para não fazer nenhum esforço durante sua caminhada até o laboratório. Será conduzido ao laboratório será apresentado ao termo de livre esclarecido e responderá a alguns questionários definidos pelo grupo. A ordem será a seguinte:

1. Termo Livre esclarecido
2. Ficha de dados criada pelo grupo
3. Estadiamento da doença nas escalas H-Y
4. Coleta de Sinais eletromiográficos.

Eletromiografia de Superfície

A eletromiografia de superfície é uma técnica não invasiva que permite o registro dos sinais elétricos gerados pelas células musculares, possibilitando a análise da atividade muscular durante o movimento. (Ocarino, Juliana de Melo, et al., 2005). No nosso estudo, os músculos dos quais serão coletados os sinais eletromiográficos são: extensor radial do longo do carpo, flexor superficial dos dedos, bíceps braquial e flexor radial do carpo. A partir destes grupos musculares pretende-se coletar sinais de diferentes tipos de tremores, como: tremor de repouso, tremor isométrico, tremor de ação e tremor de intenção.

O aparelho de sEMG utilizado foi o Miotool da empresa Miotec junto ao software MiotecSuite 1.0 para a análise dos sinais coletados. Figura 1 podemos observar o aparelho.



Figura 1: Aparelho Miotool utilizado na pesquisa (Miotec, 2016)

Para o posicionamento dos eletrodos foi protocolado que o paciente deve estar em posição de sedestação ao invés de decúbito, pois essa posição mantém o tronco ereto, tendo em vista que o paciente passará a maior parte do encontro posterior nessa posição. Os eletrodos serão posicionados acima do ponto motor.

Dados do equipamento

- Modelo do equipamento: Miotool;
- Resolução: 14 bits;

- Máxima taxa de amostragem: 2.000 amostras por segundo;
- Ruído ± 2 LSB;
- Modo de rejeição comum de 126 db;
- Isolamento de segurança 3000 V(rms);
- Tamanho aproximado de 135 mm X 140 mm X 50 mm;
- Peso aproximado de 470g;
- Tensão de alimentação do sistema de aquisição: 1 bateria NiMH 7,2 Vcc 1700 mAh;
- Corrente em repouso: 200 microA;
- Corrente máxima: 120 mA;
- Potência máxima: 0,3 W;
- Temperatura: de 10 °C a 40 °C;
- Tensão de alimentação dos canais analógicos: 3,3 V;
- Número de canais: 2 ou 4;
- Tensão máxima de entrada: 2.048 mV.
- Modelo do sensor: sDS 500;
- Tensão de alimentação: 5,0 V;
- Tensão máxima de entrada: entre 1 mV para ganho 2.000 e 8 mV para ganho 250;
- Impedância de entrada: 10¹⁰ Ohm // 2 pF;
- Ganho automático;
- Comunicação: LIN a 9.600 bauds;
- Temperatura: de 10 °C a 40 °C;
- Filtro ativo passa-baixa de dois polos com frequência de corte em 1 kHz elimina a frequências altas indesejadas;
- Comprimento do Cabo: 2 metros;
- Conexão com os Eletrodos por pressão: Cabo MiniPinch 15 cm de comprimento Blindado.

Preparação da coleta

Preparação do voluntário:

- Solicitar que o voluntário se posicione em sedestação;
- Realizar a tricotomia utilizando aparelho de barbear descartável, algodão e álcool 70%;

Posicionamento dos eletrodos:

- Eletrodo Referência: Posicionado na Crista Ilíaca direita (será utilizado 1 eletrodo por voluntário);
- Eletrodos de superfície: Deverão ser posicionados a dois dedos abaixo do ponto motor (serão utilizados 8 eletrodos por voluntário);
- Distância Intereletrodo: SENIAM: 20 mm (Serão utilizados eletrodos duplos com distância fixa);

Coleta 1 - Tremor de Repouso

Será solicitado que o voluntário permanece em sedestação com as mãos sobre o colo de forma relaxada. Serão realizadas três coletas com duração de 5 segundos cada de forma bilateral e com a utilização de 4 canais. Os grupos musculares utilizados nesta atividade são:

- Extensor radial do longo do carpo direito;
- Flexor superficial dos dedos direito;
- Extensor radial do longo do carpo esquerdo;
- Flexor superficial dos dedos esquerdo;

Na Figura 2 podemos observar a disposição dos eletrodos nas marcações em verde.

Coleta 2 - Tremor Isométrico

Será solicitado que o voluntário permanece em sedestação com as mãos cerradas sobre o colo. Serão realizadas três coletas com duração de 5 segundos cada de forma bilateral e com a utilização de 4 canais. Os grupos musculares utilizados nesta atividade são:

- Extensor radial do longo do carpo direito;
- Flexor superficial dos dedos direito;
- Extensor radial do longo do carpo esquerdo;
- Flexor superficial dos dedos esquerdo;

Na Figura 3 podemos observar a disposição dos eletrodos nas marcações em verde.

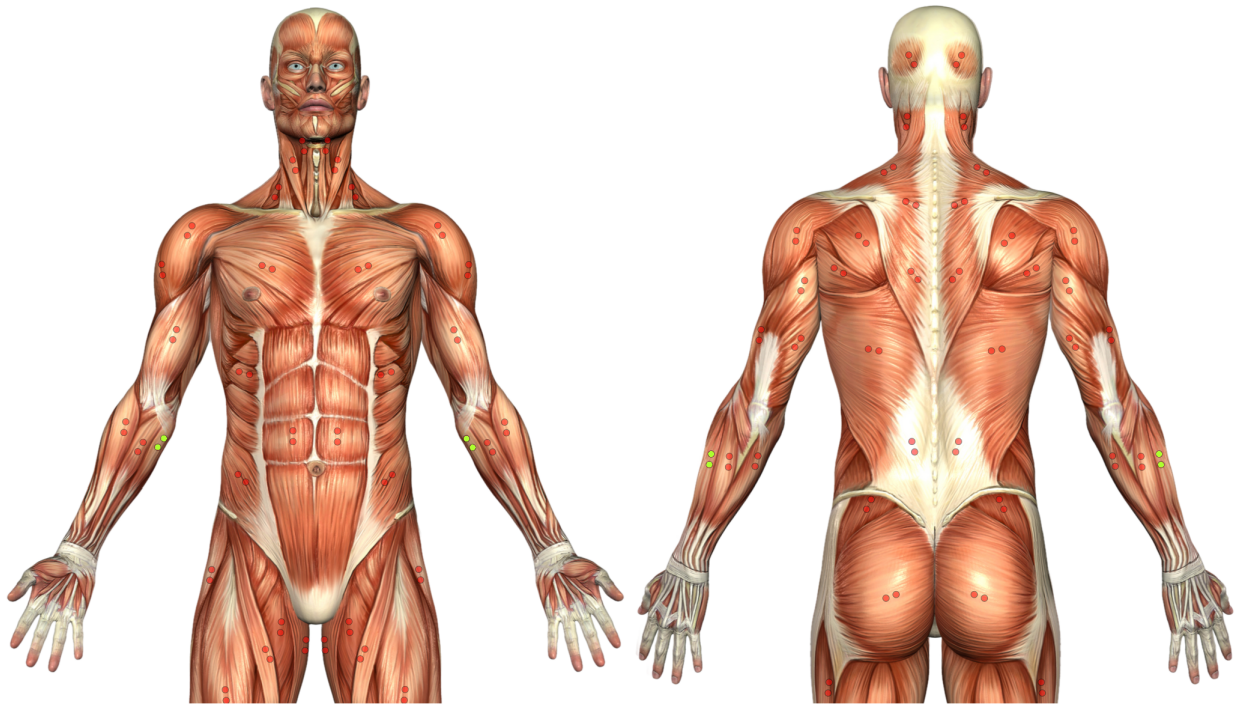


Figura 2: Posicionamento de eletrodos para captação de tremores de repouso (Miotec, 2016)

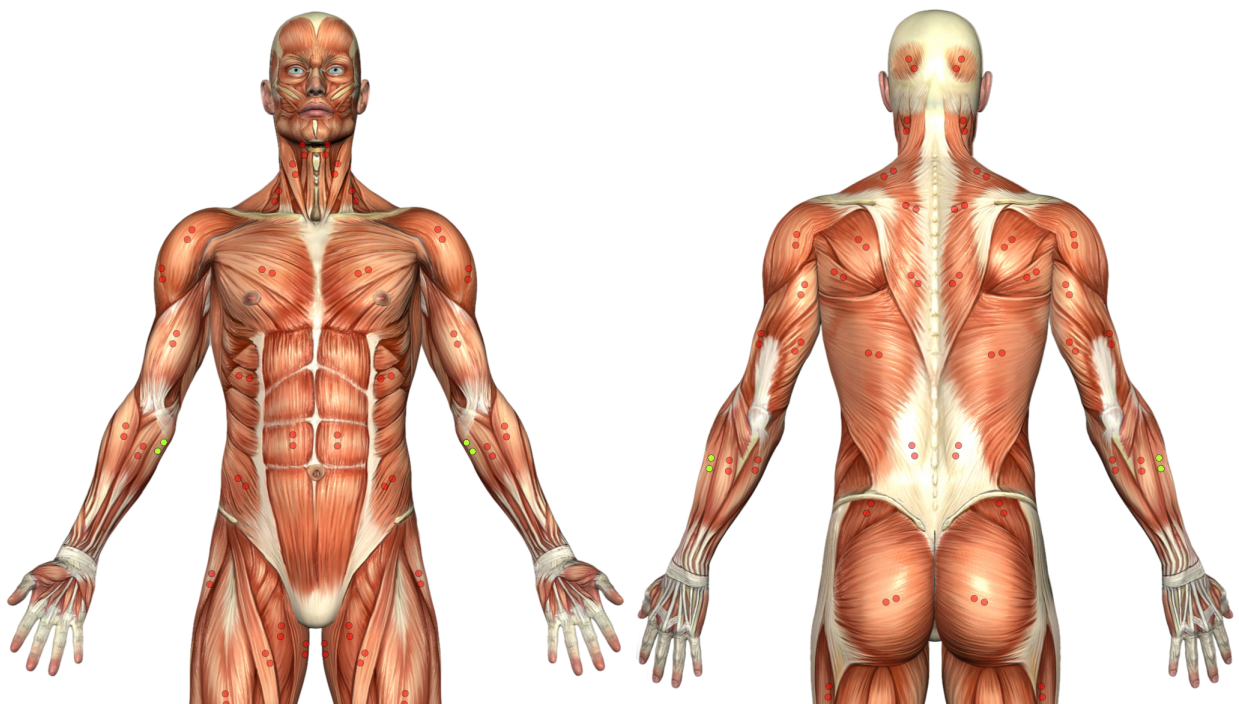


Figura 3: Posicionamento de eletrodos para captação de tremores isométricos (Miotec, 2016)

Coleta 3 - Tremor de Ação

Será solicitado que o voluntário permanece em sedestação com as mãos sobre o colo. Uma garrafa vazia sera entregue ao voluntário e então o voluntário devera realizar o movimento

de levar a garrafa até a boca e retornar ao colo 3 vezes. Este processo deverá ser repedido 3 vezes com cada um dos lados (de forma unilateral) e com a utilização de 3 canais. Os grupos musculares utilizados nesta atividade são:

- Extensor radial do longo do carpo direito;
- Flexor superficial dos dedos direito;
- Bíceps braquial direito;
- Extensor radial do longo do carpo esquerdo;
- Flexor superficial dos dedos esquerdo;
- Bíceps braquial esquerdo;

Na Figura 4 podemos observar a disposição dos eletrodos nas marcações em verde.

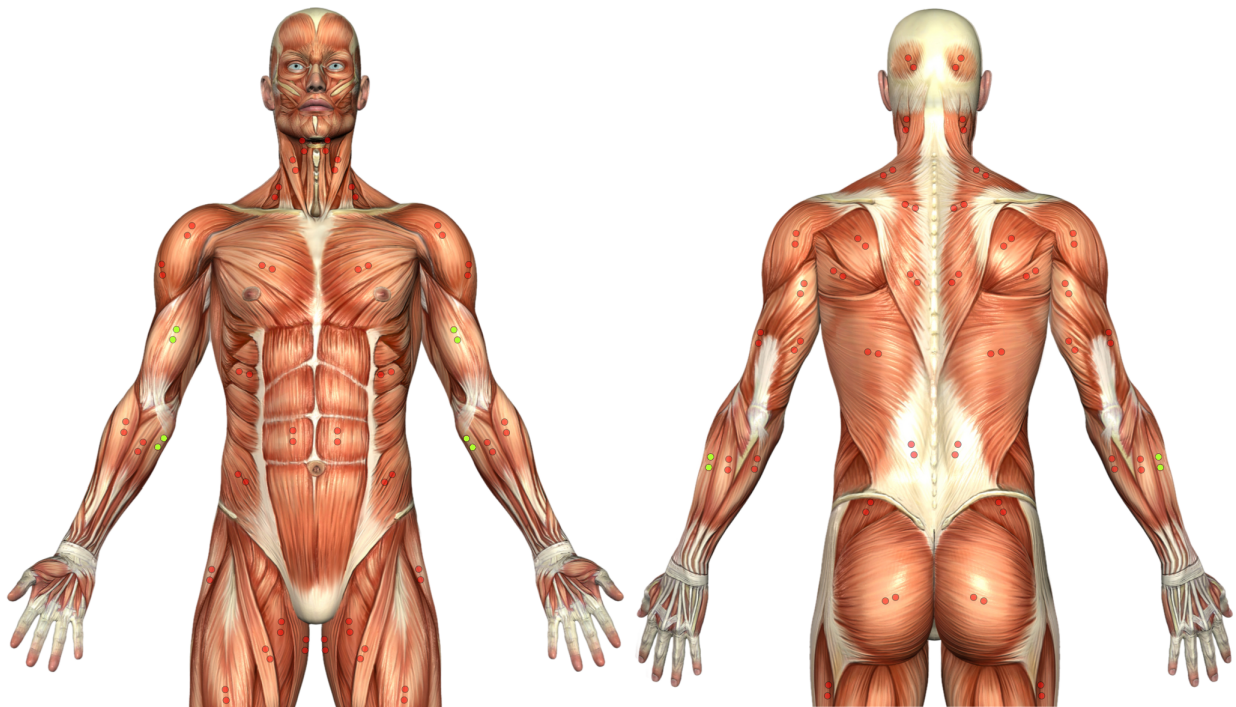


Figura 4: Posicionamento de eletrodos para captação de tremores de ação (Miotec, 2016)

Coleta 4 - Tremor de Intenção

Será solicitado que o voluntário permaneça em sedestação com o ombro em abdução de 90 graus e dedo indicador estendido. Então será instruído ao voluntário que ele alcance o seu nariz com o seu dedo indicador e em seguida retorne sua mão para a posição inicial 3 vezes. Este processo deverá ser repedido 3 vezes com cada um dos lados (de forma unilateral) e com a utilização de 4 canais. Os grupos musculares utilizados nesta atividade são:

- Extensor radial do longo do carpo direito;

- Flexor superficial dos dedos direito;
- Bíceps braquial direito;
- Flexor radial do carpo direito;
- Extensor radial do longo do carpo esquerdo;
- Flexor superficial dos dedos esquerdo;
- Bíceps braquial esquerdo;
- Flexor radial do carpo esquerdo

Na Figura 5 podemos observar a disposição dos eletrodos nas marcações em verde.

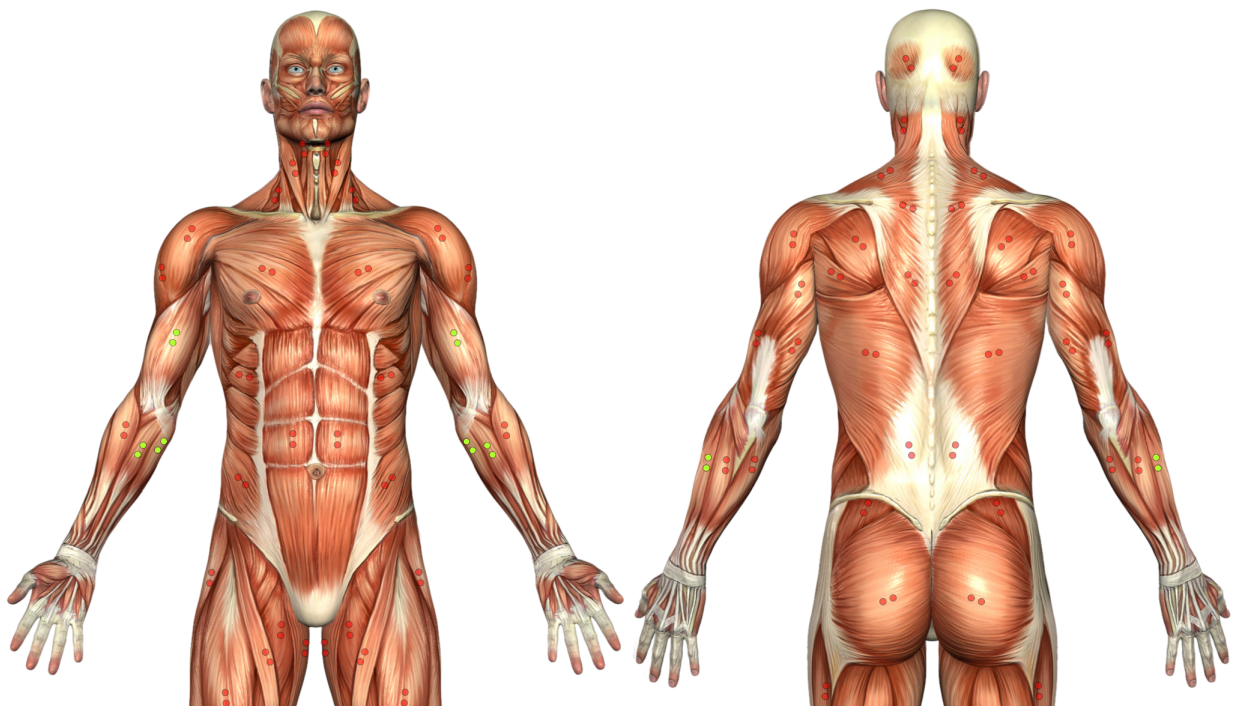


Figura 5: Posicionamento de eletrodos para captação de tremores de intenção (Miotec, 2016)

Término do encontro

Então o encontro chega ao fim, sendo oferecido um tempo de repouso e lanche para o voluntário participante da pesquisa.

ANEXO B – TAE



Termo de Consentimento Livre e Esclarecido – TCLE

O (a) Senhor (a) está sendo convidado (a) a participar do projeto: Estudo sobre o sistema cardiorrespiratório e muscular de indivíduos acometidos pela doença de Parkinson. O objetivo desta pesquisa é: Verificar possíveis complicações cardiovasculares, respiratórias e musculares causadas pela Doença de Parkinson e analisar o impacto sobre a qualidade de vida.

O (a) senhor (a) receberá todos os esclarecimentos necessários antes e no decorrer da pesquisa e lhe asseguramos que seu nome não aparecerá sendo mantido o mais rigoroso sigilo através da omissão total de quaisquer informações que permitam identifica-lo (a).

A sua participação será através de uma avaliação, a ser realizada no dia e hora que o senhor (a) puder. Informamos que o (a) senhor (a) pode se recusar a responder qualquer questão que lhe traga constrangimento, podendo desistir de participar da pesquisa em qualquer momento sem nenhum prejuízo para o (a) senhor (a).

As etapas que o senhor (a) se submeterá são:

- 1. Ficha de Avaliação contemplando nível de estadiamento da doença (HY) (aproximadamente 05 minutos);**
- 2. Eletromiografia (aproximadamente 20 minutos);**

Essa pesquisa apresenta riscos mínimos aos seus participantes, pois não serão realizados procedimentos invasivos ou nocivos a sua saúde, de forma que os riscos são: alergia aos eletrodos. Sendo que para minimizar esses riscos os pesquisadores estarão treinados para posicionamento dos eletrodos. Se você aceitar participar, estará contribuindo para a detecção de complicações da doença de Parkinson precocemente, e uma melhor na qualidade de vida dos portadores da doença. Sua participação é voluntária, isto é, não há pagamento por sua colaboração.

Todas as despesas que você tiver relacionadas diretamente ao projeto de pesquisa (tais como, alimentação no local da pesquisa ou exames para realização da pesquisa) serão cobertas pelo pesquisador responsável.

Caso haja algum dano direto ou indireto decorrente de sua participação na pesquisa, você poderá ser indenizado, obedecendo-se as disposições legais vigentes no Brasil. Os resultados da pesquisa serão divulgados na Instituição Universidade de Brasília – UnB podendo ser publicados posteriormente. Os dados e materiais utilizados na pesquisa ficarão sobre guarda do pesquisador.

Se o (a) senhor (a) tiver qualquer dúvida em relação a pesquisa, por favor telefone para: Dra. Vera Regina, na instituição Faculdade da Ceilândia, 3377 – 0615, no período da tarde, ou através do e-mail veraregina@unb.br, podendo ligar a cobrar no telefone celular 8245-5298.

Este projeto foi aprovado pelo Comitê de Ética em Pesquisa da Faculdade de Ciências da Saúde (CEP/FS) da Universidade de Brasília. O CEP é composto por profissionais de diferentes áreas cuja função é defender os interesses dos participantes da pesquisa em sua integridade e dignidade e contribuir no desenvolvimento da pesquisa dentro de padrões éticos. As dúvidas com relação à assinatura do TCLE ou os direitos do participante da pesquisa podem ser esclarecidas pelo telefone (61) 3107-1947 ou do e-mail cepfs@unb.br ou cepfsunb@gmail.com, horário de atendimento de 10:00hs às 12:00hs e de 13:30hs às 15:30hs, de segunda a sexta-feira. O CEP/FS se localiza na Faculdade de Ciências da Saúde, Campus Universitário Darcy Ribeiro, Universidade de Brasília, Asa Norte. Este documento foi elaborado em duas vias, uma ficará com o pesquisador responsável e a outra com o sujeito da pesquisa.

Nome e Assinatura

Pesquisador Responsável